LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Text-to-Speech Synthesis in Computer-Assisted Language Learning for Irish: Development and Evaluation

Neasa Ní Chiaráin

Volume 1

Thesis submitted for the degree of Doctor of Philosophy

School of Linguistic, Speech and Communication Sciences

Trinity College, Dublin

2014

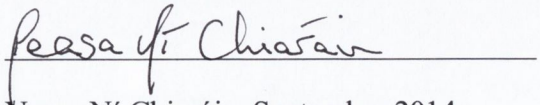
TRINITY COLLEGE
5 JAN 2015
LIBRARY DUBLIN

Thesis 10457.1

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.



Neasa Ní Chiaráin, September 2014

Summary

This thesis sets out to develop and evaluate speech synthesis-based computer-assisted language learning (CALL) platforms. Central to this work is the evaluation of the fitness-for-purpose, as part of a language teaching/learning facility, of the text-to-speech (TTS) Irish language synthesis system that these platforms employ. The TTS system in question has been developed by the *ABAIR* project at Trinity College, Dublin and is being extended and improved on an ongoing basis. There is an increasing demand for education-related applications and this was a major motivation for the present study. TTS refers to the conversion of written text into a human-like voice by a computer. CALL is a relatively new multidisciplinary area of research and it is from a task-based CALL perspective that the study is being conducted.

As a background to the study there is a general review of the literature pertaining to CALL, and discussion on the potential value of computer game-like activities as aids to second language acquisition. Essential features of effective CALL platforms are identified. TTS synthetic voices are best judged in the context in which they are likely to be used. As there were no existing CALL platforms available for Irish, it was necessary to construct prototype platforms where synthetic voices could be tested.

The criteria for evaluating CALL platforms and for evaluating the quality, intelligibility and attractiveness of the TTS synthesised voices in these platforms were identified. It is suggested that one's evaluation of any one aspect of the platforms would be influenced by the overall experience of playing the games. Thus, as one's judgment of the synthetic speech could be influenced by one's overall impression of the platforms, and might vary with the context in which one hears it, evaluations were carried out on several dimensions of the platforms in tandem with the evaluations of the synthetic speech.

Four groups carried out evaluations: teachers of Irish; 16-17 year old pupils from English-medium schools; from Irish-medium schools (Gaelscoileanna); and from Gaeltacht schools. Teachers (N=31) orthographically transcribed 20 sentences and reported on the 'ease' with which they could do so. A formula was then devised which produced a *Clarity Index*. Their opinions on the quality and attractiveness of the synthetic speech were also elicited, via questionnaires. Pupils (N=252) were given a trial session with each of the platforms. A pre-session questionnaire elicited participants' background information. A post-session questionnaire elicited their opinions on several aspects of the platforms, including the quality, intelligibility and attractiveness of the synthetic voices as well as several of the non-linguistic features of the platforms.

The teachers' responses showed that the synthetic voices were highly intelligible: transcription of the sentences was near perfect. Nonetheless, they reported a lack of 'ease' with the task indicating that factors other than 'intelligibility' play a role. While pupils' responses varied, they were generally positive about the platforms and the synthetic voices. An inferential statistical analysis showed that pupils' reactions varied with certain background features including school type, Irish language ability, frequency playing computer games, and prior attitude to synthetic speech. Overall, it was concluded that the *abair.ie* synthetic voices are fit-for-purpose for CALL game-like activities.

Having suitable synthetic speech available allows for the development of CALL platforms which can be highly interactive and personalised. A proposal for the future development of a CALL environment, in which TTS has an intrinsic part, is presented, where the strongest features of the present three platforms might be incorporated and where development is guided in equal part by the following key factors:

(a) theoretical considerations of one's views of how language is acquired
(b) the activities one organises for learners, which are consistent with these views
(c) the available technology resources
(d) the socio-cultural position and cognitive styles of the learners

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor, Professor Ailbhe Ní Chasaide, for the opportunities, guidance and encouragement she has given me throughout this process. Her dedication to advancing linguistic technology-based resources for the Irish language is exemplary and an inspiration to all of us who are involved in the field. I am grateful to Foras na Gaeilge, COGG and An Roinn Ealaíon, Oidhreachta & Gaeltachta, whose funding of the *ABAIR* project made this work possible. Partial funding for some of this work from SFI/CNGL is also gratefully acknowledged.

My sincere thanks go to Dr. Neil Peirce and Professor Vincent Wade of the Knowledge and Data Engineering Group and to Dr. Cathy Ennis and Professor Carol O'Sullivan of the Graphics, Vision and Visualisation Lab in Trinity College, Dublin, who made their software available to me to use as the platforms on which to base my research.

Special thanks to the teachers and pupils who gave of their valuable time and effort to participate in this research. Their enthusiasm for the project gave me great encouragement!

I would also like to thank my fellow doctoral students and researchers in the Phonetics and Speech Lab, Trinity College, Dublin, for their friendship and support: John Dalton, Emer Gilmartin, John Kane, Christoph Wendler, Irena Yanushevskaya, and special wishes to Amelie Dorn and Maria O'Reilly who have been a great source of support as we shared the highs and lows of 'thesis club' in the run up to our shared submission date! I would also like to give a particular mention to Harald Berthelsen for all the help he has given me throughout this process. His assistance both on the technical side as well as the encouragement he has given me have been invaluable.

The informal support and encouragement of friends and family have been indispensible, and I would like particularly to acknowledge my brother, Ruadhán, and my wonderful aunts, uncles & cousins who have all helped me reach the stage I am at today.

I am deeply indebted to my parents, Máire & Gearóid Ó Ciaráin, who have given me, all through my life, every opportunity I could ask for. It is undoubtedly thanks to my parents that my love of languages and linguistics was formed and nurtured. Ba mhaith liom buíochas ó chroí a ghabháil libh as bhur ngrá is as bhur dtacaíocht. Buíochas speisialta leatsa, Daid - ní bheadh an obair seo indéanta gan tú!

And, finally, a special thank you to Rich, for your limitless encouragement, patience and support - I am very grateful.

Contents

Page

Chapter 1: Introduction 1

1.1 Research Aim **1**
1.2 The Context of the Present Project **2**
1.2.1 Motivation. 2
1.2.2 Background. 4
1.2.3 Rationale. 7
1.3 Structure of the Thesis **9**
1.4 Definition of Key Terms **15**

Chapter 2: The Need for Speech Interactive Task-based Games for Teaching Irish as a Second Language 17

2.1 Overview **17**
2.2 The Position of the Irish Language **18**
2.2.1 Introduction. 18
2.2.2 The socio-cultural position of Irish. 18
2.2.3 Historical perspective. 19
2.2.4 The current position of the language. 20
2.2.5 Irish in schools. 23
2.2.6 The Irish language syllabus for second level schools. 25
2.2.7 The use of technology in Irish language teaching/learning. 27
2.2.8 Irish in the context of CALL for endangered languages. 28
2.2.9 Summary. 30
2.3 CALL and the Interactionist model of SLA **31**
2.3.1 Overview. 31
2.3.2 Computer-Assisted Language Learning. 32
2.3.3 Language Acquisition Theories and CALL. 35
2.3.3.1 Early approaches to CALL. 35
2.3.3.2 Communicative approaches to language acquisition and CALL. 36
2.3.4 Summary. 41
2.4 Task-based Language Learning and Teaching **42**
2.4.1 Overview. 42
2.4.2 Recent developments in communicative language teaching and CALL. 42
2.4.3 Summary. 44
2.5 Game-based Language Learning **45**
2.5.1 Overview. 45
2.5.2. Background: computer games and learning. 46
2.5.2.1 Entertainment or 'serious' games. 48
2.5.3 SLA in virtual worlds. 49
2.5.4 SLA and chatbots. 51
2.5.5 Characteristics of CALL games. 53
2.5.6 Features of CALL game design. 55
2.5.6.1 Cognitive load theory. 56

2.5.6.2 CLT and computer games 57
2.5.7 Criteria for the design of CALL games. 59
2.5.8 Summary. 60
2.6 Text-to-speech Synthesis 62
2.6.1 Overview. 62
2.6.2 Text-to-speech in CALL. 62
2.6.3 Developing a TTS synthesis system for Irish: the *ABAIR* project. 66
2.6.3.1 Unit-selection synthesis. 68
2.6.3.2 HTS synthesis. 69
2.6.4 Building *ABAIR* voices. 71
2.6.5 Currently available resources. 73
2.6.5.1 Phonetisation tool. 73
2.6.5.2 Text normalisation tool. 74
2.6.5.3 Part-of-speech tagger. 74
2.6.6. Summary. 75
2.7 Summary 76

Chapter 3: Research Design 81

3.1 Overview 81
3.2 The Development of the Three CALL Platforms 81
3.2.1 Overview of *Digichaint*. 84
3.2.1.1 Original platform: The Language Trap. 84
3.2.1.2 Adaptation: Digichaint. 85
3.2.1.3 Development of Digichaint. 90
3.2.2 Overview of *Fáilte go TCD*. 102
3.2.2.1 Development of Fáilte go TCD. 105
3.2.3 Overview of *Taidhgín*. 106
3.2.3.1 Development of Taidhgín. 107
3.2.4 Overall impact of the three platforms. 114
3.3 Evaluation Criteria 116
3.3.1 Towards an evaluation of CALL. 116
3.3.1.1 A framework for evaluating CALL. 118
3.3.1.1.1 Judgmental analysis. 118
3.3.1.1.2 Empirical analysis. 119
3.3.1.1.3 The validity and reliability of testing in the SLA/CALL context. 120
3.3.1.1.4 Chapelle's model of evaluation. 124
3.3.1.2 Criteria for evaluating the present CALL platforms. 126
3.3.2 Towards an evaluation of TTS. 128
3.3.2.1 The evaluation of TTS synthesis in a CALL context. 132
3.3.2.1.1 The influence of context on TTS evaluations. 134
3.3.2.2 Criteria for evaluating the Irish TTS synthesis in the present CALL platforms 135
3.3.3 Evaluation and evaluators. 138
3.3.4 Statistical analysis. 141
3.3.4.1 Overview. 141
3.3.4.2 Likert scale. 142
3.3.4.3 Non-parametric data. 143
3.3.4.4 Statistical analysis tests for the present study. 144
3.3.4.4.1 The Kruskal-Wallis test. 144
3.3.4.4.2 Spearman's Rank Correlation Coefficient. 146
3.3.5 Summary of evaluation criteria. 147
3.4 Statement of Hypotheses 149
3.5 Summary 151

Chapter 4: Findings for Teachers 155

4.1 Overview **155**
4.2 Participants **158**
4.3 Teachers: Research Design and Strategy **159**
4.3.1 Research Design 159
4.3.2 Research Strategy. 160
4.3.2.1 Pre-test questionnaire. 160
4.3.2.2 Test. 161
4.3.2.2.1 Sentence analysis task: evaluating 'intelligibility' and 'ease'. 161
4.3.2.2.2 Developing a Clarity Index for synthetic speech. 165
4.3.2.3 Post-test questionnaire. 166
4.4 Instruments **167**
4.5 Procedure **169**
4.6 Analysis **170**
4.7 Reliability and Validity **171**
4.8 Ethics **172**
4.9 Results **173**
4.9.1 Overview. 173
4.9.2 Teacher pre-test questionnaire findings. 174
4.9.3 Test findings. 178
4.9.3.1 Sentence analysis findings 178
4.9.3.2 Clarity Index findings 182
4.9.4 Teacher post-test questionnaire findings. 185
4.9.5 Summary of findings. 188
4.10 Summary and Relating Teacher Findings to Hypotheses **189**

Chapter 5: Findings for Learners 193

5.1 Overview **193**
5.2 Participants **195**
5.3 Learners: Research Design and Strategy **201**
5.3.1 Research Design 201
5.3.2 Research Strategy. 202
5.3.2.1. Pre-game questionnaire 205
5.3.2.2. Post-game questionnaire. 206
5.3.2.2.1 Eliciting pupils' opinions on the TTS voices. 206
5.3.2.2.2 Eliciting pupils' opinions on the CALL materials. 207
5.4 Instruments **207**
5.5 Procedure **220**
5.6 Analysis **222**
5.7 Reliability and Validity **224**
5.8 Ethics **225**
5.9 Results **227**
5.9.1 Overview. 227
5.9.2 Learner pre-game questionnaire. 229
5.9.3 Learner post-game questionnaire. 234
5.9.3.1 Overview. 234
5.9.3.2 Post-game descriptive results (non-TTS and TTS aspects). 235
5.9.3.2.1 Non-TTS descriptive results: Digichaint 235
5.9.3.2.2 Non-TTS descriptive results: Fáilte go TCD. 236
5.9.3.2.3 Non-TTS descriptive results: Taidhgín. 237
5.9.3.2.4 TTS descriptive results for the three platforms. 237

5.9.3.2.5 Summary of descriptive analysis. 239
5.9.3.3 Post-game inferential statistical results. 240
5.9.3.3.1 Overview. 240
5.9.3.3.2 Digichaint statistical analysis. 243
5.9.3.3.3 Fáilte go TCD statistical analysis. 250
5.9.3.3.4 Taidhgín statistical analysis. 255
5.9.3.3.5 Summary of inferential statistical analysis. 262
5.9.3.3.6 A holistic experience? Results of the Spearman's rho analysis. 263
5.9.4 Implications for hypotheses 265
5.9.4.1 Hypothesis 1. 266
5.9.4.2 Hypothesis 2. 270
5.9.4.3 Hypothesis 3. 274
5.9.4.4 Hypothesis 4. 275
5.9.4.5 Hypothesis 5. 275
5.9.4.6 Hypothesis 6. 276
5.9.5 Results: summary and discussion. 279
5.9.5.1 Descriptive results: summary and discussion. 279
5.9.5.2 Inferential results: summary and discussion. 280

Chapter 6: Conclusions and Future Outlook 285

6.1 Background Assumptions of the Research 287
6.2 The Development of the Research Platforms 288
6.3 The Synthetic Voices 293
6.4 Testing and Analysis 294
6.5 Summary of Teachers' Reactions to Synthetic Speech 297
6.6 Summary of Learners' Responses 298
6.7 Learners' Backgrounds and their Responses 301
6.8 Summary of Hypotheses Findings 304
Hypothesis 1 304
Hypothesis 2 304
Hypothesis 3 305
Hypothesis 4 306
Hypothesis 5 306
Hypothesis 6 307
6.9 Towards a Framework for the Development of a CALL Environment 308

References 315

Appendix A 347

A1: Digichaint Development 347
A2: Fáilte go TCD 348
A3: Taidhgín 350

Appendix B 356

B1: Teacher Pre-Test Questionnaire 356
B2: Teacher Test 358
B3: Sentences used in Teacher Evaluation 359
B4: Teacher Post-Test Questionnaire 361

Appendix C 362

C1: Pupil Pre-Game Questionnaire 362
C1.1: Irish Language Version: 362
C1.2: English Language Version: 364
C2: Pupil Post-Game Questionnaire: Digichaint 367
C3: Pupil Post-Game Questionnaire: Fáilte go TCD 372
C4: Pupil Post-Game Questionnaire: Taidhgín 376

Appendix D 380

D1: Teacher Evaluation Results 380
D2: Calculation of Intelligibility and Clarity Index 382

Appendix E 384

E1: Post-Game Questionnaire Results (Descriptive Results): Digichaint 384
E2: Post-Game Questionnaire Results (Descriptive Results): Fáilte go TCD 389
E3: Post-Game Questionnaire Results (Descriptive Results): Taidhgín 392

Appendix F 395

F1: Kruskal-Wallis Test Results: Digichaint 395
F2: Kruskal-Wallis Test Results: Fáilte go TCD 417
F3: Kruskal-Wallis Test Results: Taidhgín 440

Appendix G 456

G1: Spearman's Rank Correlation Coefficient for Digichaint 456
G2: Spearman's Rank Correlation Coefficient for Fáilte go TCD 459
G3: Spearman's Rank Correlation Coefficient for Taidhgín 461

Appendix H 463

H1: Digichaint Statistical Analysis 463
H2: Fáilte go TCD Statistical Analysis 473
H3: Taidhgín Statistical Analysis 478

Chapter 1: Introduction

1.1 Research Aim

The aim of the present study is to develop and evaluate Irish language computer-assisted language learning (CALL) platforms which have integrated synthetic speech as a key feature. As part of this, it specifically aims to evaluate the intelligibility, quality and attractiveness of synthetic Irish language voices which are an integral part of the newly developed platforms. Irish language text-to-speech (TTS) synthesis is still at a developmental stage and the present work represents an initial evaluation of the emerging speech synthesis system, known as *ABAIR*, in the specific context of three customised CALL platforms developed for this thesis.

TTS synthesis refers to the conversion of written text to an artificial human-like voice by a computer. Such voices are frequently referred to as 'synthetic voices'. CALL refers to the use of computers as aids to the enhancement of the language learning process.

This study aims to evaluate three CALL platforms which exploit the Irish synthetic voices of *ABAIR*, using teachers and learners of Irish from second level schools in Ireland. There is an emphasis on using teachers and learners with a high Irish language comprehension ability (learners from Irish-medium schools) as evaluators so that one can have greater confidence that any weaknesses which they report in the synthetic speech are actual faults with the synthesis rather than a low level of comprehension ability amongst the language learners. Nonetheless, a significant cohort of learners from English-medium schools, for whom the CALL platforms are specifically aimed, is included. The study will seek to elicit the opinions of these teachers and learners on the design features of the CALL platforms; their value as language learning aids; and the intelligibility, quality and attractiveness of the TTS synthetic voices used.

Specific Research Objectives

The general research aim can be expanded into the following more specific set of objectives:

- To develop prototype CALL materials (i.e. game platforms) which are in harmony with current theories of second/foreign language acquisition – for the purposes of the present study these are linked to task-based language learning/teaching principles
- To evaluate the salient aspects of the platforms as aids to language learning/teaching
- In particular, to evaluate the *ABAIR* TTS synthesis system for use in such platforms, in terms of its intelligibility, quality and attractiveness (see Section 1.4 for definition of terms)
- To explore the value which the TTS synthetic speech can add to the CALL platforms as an aid to language learning/teaching
- To derive insights from the development and evaluation phases of the project in order to make proposals for the blueprints of future CALL platforms

1.2 The Context of the Present Project

1.2.1 Motivation.

Over 44,000 pupils take the Leaving Certificate state examination in Irish each year and there are approximately 700,000 learners of Irish in the education system in the Republic of Ireland. Numbers in Northern Ireland, where pupils may study Irish for varying periods during their school years, are more difficult to quantify. The latest figures that are available show that 2,999 pupils are in funded Irish-medium education settings at preschool, primary and post-primary levels (John O'Dowd, Minister for Education, Northern Ireland Assembly Debates, 20 February 2012). No more than a few hundred

take Irish in the A-level examination which would be somewhat equivalent to the standard of the Leaving Certificate in the Republic.

Beyond learners of Irish in Ireland, there is a very considerable interest in the Irish language from abroad. This is evidenced by the ever-increasing numbers visiting the *ABAIR* project website (*www.abair.ie*) where the recently developed TTS facility for two Irish dialects can be accessed by the public. More than half the visits to the *www.abair.ie* website come from North America and the remainder come from divergent places across the globe ranging from South America to Asia, Australia and the Middle East. For learners and for those curious about the language, the *ABAIR* synthetic voices offer the only possibility of readily matching any written text to its pronunciation - a valuable resource particularly for those who have no access to native speakers.

Returning to the situation in Ireland, the technological facilities available for the teaching of Irish in schools are very limited at present. While access to the spoken language is widely available from the media as well as from specifically designed teaching programmes, most learners do not readily have access to native speakers of the language. The possibility of interactive speech output software would open up entirely new opportunities for the teaching of Irish. For the first time, learners could input their unique materials and have them read aloud in native human-like speech. The context in which Irish is taught is somewhat unique and qualitatively different from that of the major foreign languages taught in Ireland. Being a less widely spoken language, the range of possible target language experiences which the learner can have is restricted, given the limited opportunities to interact with a native speaker of Irish in the target language. Synthetic speech, when integrated into a suitable CALL platform, would offer the learner the possibility of interacting in meaningful ways with native human-like voices. It could also make learning a more interactive and enjoyable experience and possibly increase learner motivation. The present situation and status of the Irish language is discussed in detail in Chapter 2.2.

The use of synthetic speech in CALL is at a tentative stage in its development. The understandings gleaned from the present project have applications which go beyond the teaching of Irish and into the realm of second/foreign language learning in general. The opportunities offered by new speech technologies, which can allow for a more active participatory role for the learner, have a significant contribution to make to our understanding of the sociocognitive processes involved in language learning, i.e. the interaction between context or learning environment and learners' desire and ability to acquire the target language.

For the present research three platforms have been developed to the level of proof-of-concept. They are tested on teachers and learners in order to gauge their reactions to various facets of the platforms. A framework is proposed on to which future CALL developments can be added as technological developments proceed, including imminent developments in the quality of synthetic speech. Such a framework would ideally be open-ended and have the facility to allow non-expert computer users to develop new storylines and new activities. Such materials could add greatly to the motivation of learners and make the learning process a more enjoyable experience.

In summary, the present study aims to develop suitable CALL platforms for learners of Irish in second level schools, to integrate TTS speech into these platforms, and to evaluate both the TTS and non-TTS aspects of the platforms as an aid to Irish language learning.

1.2.2 Background.

Work on the development of synthetic speech has been underway for the past 30 to 40 years and speech synthesisers have been included in many computer operating systems since as early as the 1980s. Considerable progress has been made to date in the

development of TTS and robust systems with high quality human-like voices have been developed for the major world languages.

The *ABAIR* project is the first significant development in producing synthetic speech for Irish. The Irish language has a considerable dialectal variation, with three main dialects. Unlike many other languages, Irish has no accepted standard spoken dialect. Proposals for standardisation of the forms of written Irish, which draw on and combine forms from the main dialects, have been made (see *Lárchanúint don Ghaeilge* (Ó Baoill, 1987)). Despite the possibilities that it offered, the standardised dialect has not gained widespread support in the media or in education, and there are no speakers of this 'Lárchanúint'. For those not familiar with modern synthesis techniques, it is perhaps worth pointing out that the two dominant approaches to TTS building, "unit-selection synthesis" and "HTS" (used for the ABAIR voices, see Section 2.6.3) are built on the basis of a spoken corpus, recorded for a speaker of the language/variety. As mentioned earlier, there are no speakers of Lárchanúint and its principle value is that is provides a written standard. Even should one have wanted to develop a rule-based TTS system using formant synthesis, it would effectively require extensive analysis of real speech by a speaker of the language/variety. Given these considerations, a lárchanúint TTS was not really on the cards. Even if one developed such a system, it is not obvious that its output would be acceptable to listeners of Irish. Thus, from the beginning of the *ABAIR* project, a multidialectal approach was adopted, and the initial goal was to create TTS synthesis for each of the three main dialects of Irish – Ulster Irish (as represented by a Gaoth Dobhair speaker), Connaught Irish (as represented by a Ráth Chairn speaker) and Munster Irish (as represented by a Corca Dhuibhne speaker). Within each dialectal area there can be some variation (Dalton & Ní Chasaide, 2007). However, as the currently prevalent synthesis systems are built on a corpus from a specific speaker, within-dialect variation is not a concern here. The work on the development of synthetic voices for Irish dialects is being carried out as part of the *ABAIR* project in the Phonetics and Speech Laboratory, Centre for Language and Communication Studies (CLCS), Trinity College, Dublin, and the voices developed to date are available online at *www.abair.ie*. As a member of the

ABAIR research team I have been actively involved in the development of synthetic speech for each of the three main dialects of the Irish language, and I have been involved in the recording and annotation of corpora, as well as in the development of some of the underpinning linguistic resources. The focus of the present study is related to the *ABAIR* project in that it examines the possible exploitation of the TTS synthesis, as developed to date, as an integral part of CALL platforms for the teaching/learning of Irish.

The use of synthetic speech for second/foreign language teaching/learning is a subject of some considerable debate among proponents of CALL (Sha, 2010). In the past, it was constantly feared that the quality of the speech in terms of its accuracy and prosodic features was such that it might give poor quality speech samples to the second/foreign language learner. However, as the quality of synthetic speech has improved greatly in the last twenty years, it is now useful to examine its possible exploitation as a teaching/learning aid. While the use of synthetic speech for various aspects of CALL has been researched by Handley (2009) and Handley and Hamel (2005) there has, to date, been no investigation of its use in games, or for computer-based activities specifically designed for language teaching/learning.

The present thesis will investigate the use to which (Irish) TTS synthesis can be put in the context of CALL games and evaluate its intelligibility, quality and attractiveness, as perceived by learners and teachers. It is recognised that the present evaluation must be distinguished from an overall evaluation of the *ABAIR* TTS synthesis system. Here, the focus is on the evaluation of the TTS voices as a facility for use in specific platforms described in this project. A further general evaluation of the overall *ABAIR* output will be necessary at a future date.

Since there were no suitable pre-existing materials on which to test the Irish synthetic speech, these had to be developed as part of the present study. Details of the adaptation and development of the platforms are given in Chapter 3.2. The main focus of this study

is the integration of Irish synthetic speech into interactive CALL platforms. This is being discussed from a CALL perspective and consequently a considerable amount of the thesis is devoted to an analysis of CALL in its broad context. As part of this analysis the place of interactive digital platforms within CALL is investigated, as is the usefulness of TTS synthesis as an integral part of these platforms.

The study of TTS synthesis within CALL crosses many potentially distinct academic disciplines, including language learning theory, linguistics, game development, speech science and social science. Each of these disciplines brings perspectives that are crucial to the overall development of specific CALL materials in the Irish language context. The relevant theoretical background is explored primarily in Chapter 2 of the present thesis.

While working on this project I am in the somewhat unique position of contributing to the development of the TTS synthesis system and simultaneously developing the CALL platforms on which the synthetic voices are tested. This has permitted me to make adjustments to the TTS system or to 'train' it specifically for the requirements of the CALL platforms being developed. The *ABAIR* synthetic speech is undergoing continuous development and the platforms developed here serve also to offer suitable contexts in which the *ABAIR* voices can be exploited and evaluated.

1.2.3 Rationale.

The rationale for the present study can be divided into five main considerations:

(a) the relative paucity of CALL resources for Irish
(b) the importance of native-speaker models being available for learners of Irish
(c) the development of interactive platforms which can facilitate personalised and self-directed learning
(d) the evaluation of the *ABAIR* TTS voices for the first time in a CALL environment
(e) the assessment of the implications of such platforms for the development of CALL in general

These considerations are expanded as follows:

(a) The world's most widely used languages invariably attract large numbers of learners from outside their linguistic communities for a variety of reasons including trade and political considerations. The teaching of such languages often represents significant industries in themselves and can be well resourced from within their own commercial organisations. The opposite applies to less widely used languages as demand for acquiring them is relatively small and resourcing the process is not commercially viable (Ward, 2004). Irish is a typical example of such a language. There are few, if any, interactive digitalised platforms available as teaching resources. The present study aims to develop prototypes of CALL platforms which could in principle be made available as teaching/learning resources for schools.

(b) CALL is a relatively new multi-disciplinary field which offers learners the possibility of personalised and self-directed learning (see Chapter 2.3 for a more in-depth discussion on CALL). Few such facilities have previously been made available for Irish. This is the first time that synthetic speech is available for inclusion as part of such a platform. The availability of native-like pronunciation of Irish words and phrases is particularly important since few will have much contact with native speakers of the language (see discussion on the current position of the Irish language in Chapter 2.2).

(c) CALL platforms are at a stage of development where they can interact with the learner in a non-linear fashion, in a way that can produce novel discourse turns. It is not possible to prerecord such utterances with human voices as had been done for earlier audio-visual language courses, since the utterances produced by the computer will be novel. The system may accept input from the learner and react to it giving appropriate output, thereby creating novel utterances. The program which the learner follows can be personalised and its direction largely determined by the learner. See Chapter 2.5.4 for a

general discussion on chatbots, or *dialogue systems*, and Chapter 3.2.3 for the development of such a chatbot which can interact with learners in a non-linear fashion, for the purposes of the present thesis.

(d) Part of the present study represents the first evaluation of the emerging Irish synthetic speech from the *ABAIR* system. It is being evaluated by teachers and by learners of Irish in game-like CALL platforms. No evaluations of synthetic speech have been done previously in any language in such a context. See Chapter 3.3 for a discussion on evaluation criteria for synthetic speech and Chapters 4.9 and 5.9 for results of teachers' and learners' evaluations of the *ABAIR* synthetic voices in the context of the newly developed CALL platforms.

(e) As well as an evaluation of the TTS synthetic voices, this study also evaluates the various design aspects of the CALL platforms themselves and their usefulness as facilitators of language learning. The evaluation of CALL platforms with integrated synthetic speech may have implications for the development of CALL in general which go beyond that of CALL for the Irish language (see Chapter 6.9).

1.3 Structure of the Thesis

The thesis is divided into 6 chapters. This section outlines the content of each chapter in turn. Section 1 of each chapter gives an overview of that specific chapter. A summary of the main findings is given at the end of each chapter.

Chapter 1 is the introductory chapter which gives the context and background from which the project emanates and states the general research aims.

Chapter 2 is broken down into a number of sections. It consists of a literature review of

each of the areas which is pertinent to the development of CALL platforms in the context of using TTS synthesis for the teaching of Irish.

Chapter 2.2 examines the position of the Irish language in Ireland at present and the attitude of the population towards it. This gives an indication of pupils' motivation to learn Irish at school. The huge variation in linguistic backgrounds of those studying Irish in school is noted, as this is highly relevant to the groups on which the present platforms have been evaluated. A brief summary is given of the use of technology in Irish language classrooms at present.

In Chapter 2.3 the role of CALL in the process of language learning/acquisition is examined in its historical context. The relationship between the emergence of CALL and prevailing philosophical, psychological and pedagogical theories is explored. The diversity of activities which fall into the general CALL category is outlined. Past and current pedagogical theories are discussed and the likely future direction of CALL development is identified. The implications of the current theory of task-based communicative language teaching for the design and development of CALL activities are identified in Chapter 2.4.

Chapter 2.5 examines computer games and the relationship between general computer gaming and games specifically designed for some educational purpose. It reviews the criteria associated with the development of various platforms which have a specific language teaching/learning function. It reviews the role to date of TTS synthesis in a CALL context and identifies the difficulties seen by researchers in using TTS synthesis for language teaching/learning. The adequacy or fitness-for-purpose of TTS in CALL is discussed.

Chapter 2.6 provides background information on the *ABAIR* Irish language synthesis system and the two different methods of TTS being developed in the

Phonetics and Speech Laboratory, TCD. The relative merits and drawbacks of each method are discussed. The section contains an outline description of the process involved in building a synthetic voice and gives a description of the linguistic resources that underpin the synthesiser.

Chapter 3 details the development and evaluation of three specific CALL platforms and gives a detailed description of the tools of evaluation including the statistical analysis to be applied.

Chapter 3.2 gives a detailed account of the development of the three multimedia tools which are being used to serve as platforms for the testing of the *ABAIR* synthetic voices. The unique characteristics of each of the three platforms are described. The first platform, *Digichaint,* was adapted so that it became a suitable vehicle to which Irish TTS voices were added. The highly sophisticated graphics and manipulation system were adapted to make them suitable for the Irish context. The most salient feature of the second platform, *Fáilte go TCD,* is its avatars and their highly stylised movements which bring them as close as possible to the movements of humans. The third platform, called *Taidhgín*, was developed from Pandorabots and is an original development. It can be described as a 'chatbot' and is programmed to accept input from learners and to act as a dialogue partner.

Chapter 3.3 considers the criteria for evaluating both aspects of the CALL platforms – the TTS and non-TTS aspects. It first reviews the general literature pertaining to the evaluations and then extracts the specific criteria which will be used for the evaluation of the present CALL platforms. It considers the suitability of different categories of evaluators and the context in which the evaluations could take place. It further outlines the statistical techniques which may be used for the analysis of the data generated from the responses of the evaluators.

In the light of the general theoretical considerations and their relevance to the present context, a set of six specific hypotheses which may be tested is formulated in Chapter 3.4 and these form the basis for the subsequent evaluations.

Having identified teachers as suitable participants in the evaluation process, Chapter 4 gives a detailed account of the teacher evaluation section of this study.

Chapter 4.2 gives details on the participants' backgrounds and the design of the evaluation process is outlined in Chapter 4.3. This process involves teachers orthographically transcribing twenty sentences spoken with the *ABAIR* synthetic voices. The methodology for evaluating the teacher transcriptions is outlined and a formula is developed which combines the accuracy with which each respondent transcribed each utterance, and the 'ease' with which they did so, as reported by themselves. The instruments used to collect the data, the procedures for doing so, the data analysis, considerations of reliability, validity and ethics are described in Chapters 4.4 – 4.8. Chapter 4.9 sets out the results and analysis of the data obtained from these evaluations of the TTS synthetic voices and their suitability in the context of CALL games. Results show that the twenty sentences were highly intelligible and teachers transcribed them with a 91% accuracy rate. This high rating, however, needed to be qualified since evaluators reported that in transcribing the sentences, they experienced difficulty in doing so. An overall score for the *Clarity Index*, which combines both the 'intelligibility' and the 'ease of intelligibility' ratings, is calculated by applying the formula developed earlier in Chapter 4.3.

Chapter 5 mirrors the structure of Chapter 4 and gives a detailed account of the learner evaluation section of this study. Unlike Chapter 4, which produces both subjective and objective findings, Chapter 5 considers the subjective responses of the pupil participants to both the TTS and non-TTS aspects of each of the three CALL platforms.

Chapter 5.2 sets out details pertaining to the backgrounds of the pupils. Justification for the design of the study is outlined in Chapter 5.3 and it was decided the data would be collected by means of Likert scale-based questionnaires. The connection between these questionnaires and the theoretical background on which they are based is tabulated in Chapter 5.4. Chapters 5.5 - 5.8 describe the procedure involved in administering the evaluations, the system of analysis used, and matters relating to reliability, validity and ethics.

The results are set out in Chapter 5.9. Chapter 5.9.2 details the learner pre-game questionnaire findings, which outline background information on the participants. The post-game results are divided into two main sections. Chapter 5.9.3.2 gives a descriptive account of the responses given by the evaluators to the questionnaires. Chapter 5.9.3.3 details the inferential statistical analysis and uses the Kruskal-Wallis test to establish the relationship between certain background characteristics pertaining to the respondents and their reactions to the platforms. The hypotheses, set out first in Chapter 3.4, are tested for statistical significance and implications for each of the hypotheses in turn are discussed in Chapter 5.9.4. It is established that one's prior disposition towards synthetic speech was one of the prime factors associated with one's evaluation of the intelligibility, quality and attractiveness of the voices. Also, one's opinions on the intelligibility of the synthetic voices were related to one's overall competence in the Irish language. Overall, the group who found the platforms most attractive were those who had an intermediate level of Irish and who saw such platforms as having a relevance to them. They were most enthusiastic about the usefulness of CALL materials to enhance the language learning process for Irish. Those who had a high language competence were less attracted to the platforms because they did not see them as having an immediate relevance to themselves. A third factor, namely the frequency with which one played computer games, was also significantly related to a number of items in the evaluation. Computer game playing was not an activity in which most of the pupils

partook very frequently. Those who never played computer games were least appreciative of the platforms. The most appreciative group were those who played games occasionally. They had sufficient skills to manipulate their way through the gaming process but were not comparing it to commercial games. Probably the most significant result was drawn from the Spearman's rho test (see Chapter 5.9.3.3.6) which showed significant correlations between almost all features (both TTS and non-TTS) of the platforms. This suggests that the games were seen as being holistic experiences and that one's opinion of any one aspect of the platforms was coloured by one's opinions of the other aspects. This leads to the proposition that one's judgment of synthetic speech is context-dependent rather than being a constant. This proposition has important implications for how we look on synthetic speech evaluation suggesting that the synthetic speech may be viewed differently by evaluators depending on the context in which they hear it.

Chapter 6 summarises the thesis and its main findings. It notes that the TTS synthesis was regarded as being of a sufficiently high standard in terms of its intelligibility and quality in the context of personalised interactive CALL platforms. Consideration is given to the implications which this finding has for the development of CALL materials in general and the new interactive dimension which TTS synthesis can bring. It opens up new possibilities for the development of communicative approaches to language teaching using digitalised, or virtual worlds. A representation of an ideal platform is given in diagrammatic form and proposals are made for the development of CALL materials (1) which can fall within this framework, (2) be compatible with each other, and (3) form an ever-evolving resource. Four modules, which were identified as having key influences on the evolution of the proposed platform are represented as '*Why?*', '*What?*', '*How?*' and '*For Whom?*'. The *Why* represents the theoretical underpinning for the approaches which are being taken to language teaching. The *What* represents the actions or activities which are compatible with the theoretical framework. The *How* represents the technological facilities which are available and the *For Whom* represents the learners who may form disparate groups in terms of their cognitive processing, the psychosocial position in

which they find themselves, and their affective disposition towards language learning. It is the interaction between these four modules that will give direction to successful future CALL learning environments.

1.4 Definition of Key Terms

Platform: refers to a system which operates in accordance with specific predefined rules and may host any number of resources which are mutually compatible. The present thesis refers to CALL platforms, which aim to engage learners with a target language in a meaningful way. There are three platforms referred to in the present study and each is at a prototype stage. This offers the possibility of expansion and the addition of new material which is compatible with the existing format. Essentially a 'platform' is being referred to as any functional piece of software which has been specifically designed to facilitate language acquisition. Platforms may be stand alone or may be interconnected to each other in the sense of sharing databases or software features.

Intelligibility: refers to the listener's ability to decode the output produced by the TTS system. Note that the term 'intelligibility' here is not being confined to a recognition of individual phonemes or words in isolation as used by a number of investigators towards the beginning of the century such as Koul (2003). Rather, it is used as the listener's ability to recognise (and orthographically transcribe) a set of sentences but not requiring any higher level cognitive functioning such as constructing a coherent mental representation of the information contained in the text and relating it to any pre-existing knowledge which the listener may have (Kintsch & van Dijk, 1978).

Quality: implies an aesthetic evaluation of the speech. It refers to system adequacy in terms of it being free from aural hindrances such as buzziness, lack of smoothness or discontinuities of speech.

Attractiveness: refers to the tendency of the synthetic voice to evoke a negative or a positive disposition within the listener. It most likely will have an important influence on the extent to which the listener may be willing to use the synthetic voice in future applications.

Clarity: refers to a combination of 'intelligibility' and 'ease of intelligibility' – a formula to produce a *Clarity Index* is newly developed as part of this thesis and is applied in Chapter 4 in the teacher evaluation section. A rating for 'intelligibility' is counted as a *performance* score and a rating for 'ease of intelligibility' is counted as an *opinion* score. The *Clarity Index* combines both the *performance* and *opinion* scores in order to produce one overall score. This *Clarity Index* is a newly developed measure which is not taken from previous research.

Objective / Subjective Data: For the purposes of this thesis, measurable objective data acquired from the teachers' transcription exercise (see Chapter 4) are being referred to as '*objective*' data. Opinion-based data acquired from both teachers and pupils through the use of Likert scale-based questionnaires (see Chapter 5) are being referred to as '*subjective*' data.

Chapter 2: The Need for Speech Interactive Task-based Games for Teaching Irish as a Second Language

2.1 Overview

This chapter outlines the present situation of the Irish language, including the strengths and weaknesses of its socio-cultural position, as it is in this context that we understand the necessity of CALL materials for Irish. The situation of Irish language teaching/learning in schools is also discussed and put in the wider context of the general practice of second/foreign language teaching/learning (see Section 2.2).

The relationship between developments in CALL historically and language learning/acquisition theories is also explored here (see Section 2.3), focusing particularly on the interactionist model of second language acquisition (SLA) to which the present thesis subscribes. This interactionist model is realised in terms of task-based learning/teaching activities, which are explored in Section 2.4. One such task-based activity is the playing of language games. The key features which add to the effectiveness of computer-based games are discussed in Section 2.5.

Finally the place of TTS synthesis in CALL is examined in Section 2.6 and a detailed description of the *ABAIR* Irish language TTS system is given. A summary is included in Section 2.7, where a brief account of the main points made in this review of the literature is given.

2.2 The Position of the Irish Language

2.2.1 Introduction.

This section begins with the current socio-cultural position of the Irish language both in society generally and in schools in particular (see Sections 2.2.2, 2.2.3 and 2.2.4). The relative strengths and weaknesses of its current position are outlined as are the factors which have led the United Nations Educational, Scientific and Cultural Organization (UNESCO) to enumerate it amongst 'definitely endangered languages' (Moseley, 2010).

The position of Irish in the school system is outlined and brief consideration is given to computer-based teaching materials which are currently in use (see Sections 2.2.5, 2.2.6 and 2.2.7). Consideration is given to the position of Irish as an 'endangered language' and to the parallel use of CALL materials in other endangered languages from various parts of the world (see Section 2.2.8). The value of digital materials for use in the teaching of Irish is discussed in terms of its motivational potential for learners.

2.2.2 The socio-cultural position of Irish.

The status of a language within society will affect people's motivation for learning it. It is essential to describe the social context in which any teaching/learning material is being developed as it may have a significant influence on the acceptability, or otherwise, of the material. Irish is in a somewhat unique position as a language which has very considerable official support, and yet, motivation to acquire the language amongst younger people would appear to be rather weak or ambivalent. The motivation for learning Irish may be quite different from that of learning other languages since there is no longer a community of monolingual Irish speakers with which the learner may wish to associate.

Irish is taught in all state-funded primary and second level schools. Linguistically, schools fall into three categories:

(a) *English-medium schools*, which teach all subjects except Irish through the medium of English. These represent the vast majority of schools throughout the country. The latest figures show there are 655 second level English-medium schools in the 26 counties (Department of Education and Skills, 2013)

(b) *Gaelscoileanna*, in which all subjects are taught to pupils through the medium of Irish. Most of these pupils have English as their home language. There are 36 second level Gaelscoileanna in the 26 counties and 5 in Northern Ireland (Gaelscoileanna, 2013)

(c) *Gaeltacht schools*, which are located in Irish speaking areas of the country. These cater both for pupils who are native speakers of Irish and for pupils whose home language is primarily English. There are 30 second level Gaeltacht schools in Ireland (Mac Donnacha, Ní Chualáin, Ní Shéaghdha, & Ní Mhainín, 2005)

2.2.3 Historical perspective.

Irish is a Celtic language, and together with Scots Gaelic and Manx traces itself to the Q-Celtic branch, which was spoken by the Celtic people who came to Ireland about 500 BC. The Q-Celtic language developed into what is termed *Old Irish*, which in turn, developed into *Scots Gaelic*, *Manx* and *Middle Irish* around the tenth century and *Modern Irish* from the middle ages on. It was the most widely spoken language on the island up to the early 19th century, but for reasons associated with the political, economic and legal situation that pertained in Ireland in the 18th and 19th centuries it went into rapid decline (O'Beirne Ranelagh, 2012). The Great Famine of 1845 had a devastating effect as the majority of its victims were Irish speakers who lived in poorer areas. At the end of the 18th century about 50% of the population of the island were monolingual speakers of Irish. In less than 100 years that followed, monolingual speakers were reduced to a mere 1% (An Coimisinéir Teanga, 2008). The cultural revival movement of the 1880s and 1890s came somewhat too late as the critical mass of monolingual speakers had declined below the threshold where a full revival was possible, particularly given the very poor economic circumstances of Irish speakers, their widespread emigration, and their lack of political power and influence.

The new Free State Government put its weight behind the revival of Irish and Article 4 of the 1922 Constitution designated Irish as the national language of the Free State. State support for the language continued and Article 8 of the 1937 Constitution reflects the importance attached by the Government to the language. It kept its Constitutional superiority over English as the first official language and Article 35 went on to state that the Irish version of the text of a law enrolled in both official languages shall prevail in the event of a linguistic discrepancy existing between the two versions (Ó Murchú, 2001). Article 8 of the Constitution states that English is recognised as the second official language, but provision may be made by law for the exclusive use of either of the two languages for any one or more official purposes, either throughout the State or in any part thereof (Bunreacht na hÉireann, 1937).

Irish was a compulsory requirement for employment in the public sector until 1974. Changes to the regulations put Irish and English on an equal footing so that candidates could use either language, or both, in entrance competitions. By the end of the 1970s it became increasingly difficult for the civil service to provide a service through Irish (An Coimisinéir Teanga, 2008).

In 2003 the Official Languages Act came into law placing on a statutory footing the right of the citizen to get services through Irish from the State. The legislation was intended to give tangible effect to the Constitutional provision. In January 2007 Irish was given official status in Europe and became the 23rd language of the European Union. It was accorded the status of a treaty language which means that all key EU legislation will be translated into Irish (Mac Giolla Chríost, 2012; Ó Riain, 2010; Walsh, 2012).

2.2.4 The current position of the language.

Despite the enormous official support for the language, UNESCO classified Irish as being in its "definitely endangered" category (Moseley, 2010). According to the 2011 Census of Population 1.77 million or 40.6% of the population of the Republic over the

age of 3 are able to speak Irish. These figures have recently been made available online (CSO, 2011). Regular speakers of Irish were mainly involved in schools and education. Figure 2.1 below shows the distribution of Irish speakers throughout the Republic of Ireland. Just 77,185, or 1.8% of the population aged 3 and over, said that they spoke Irish on a daily basis outside of the education system. The higher the education level attained by individuals the greater the chance that they could speak Irish. Only 12.4% of people who left school with primary education could speak Irish while 43.8% of those who had reached graduate level could do so. Perhaps the greatest danger to the future of the Irish language is exemplified in the statistic that only 24% (23,175 persons) of those who live in Gaeltacht areas speak Irish on a daily basis outside the education system. Some 4,682 persons in the Gaeltacht areas indicated that they could speak Irish but didn't do so on a regular basis. One needs caution in dealing with any statistics pertaining to Gaeltacht areas since official boundaries are long established and are somewhat out of date since inroads into these boundaries by an English-speaking community have been occurring almost since their establishment. The Planning and Development Act 2000 states "when making regional planning guidelines which affect the Gaeltacht, the regional authority shall have regard to the need to protect the linguistic and cultural heritage of the Gaeltacht" (Section 23: 4b). There have been some disputes in relation to how vigorously this section of the act has been applied. Certain areas of Connemara, for example, are now part of the suburbs of Galway city.

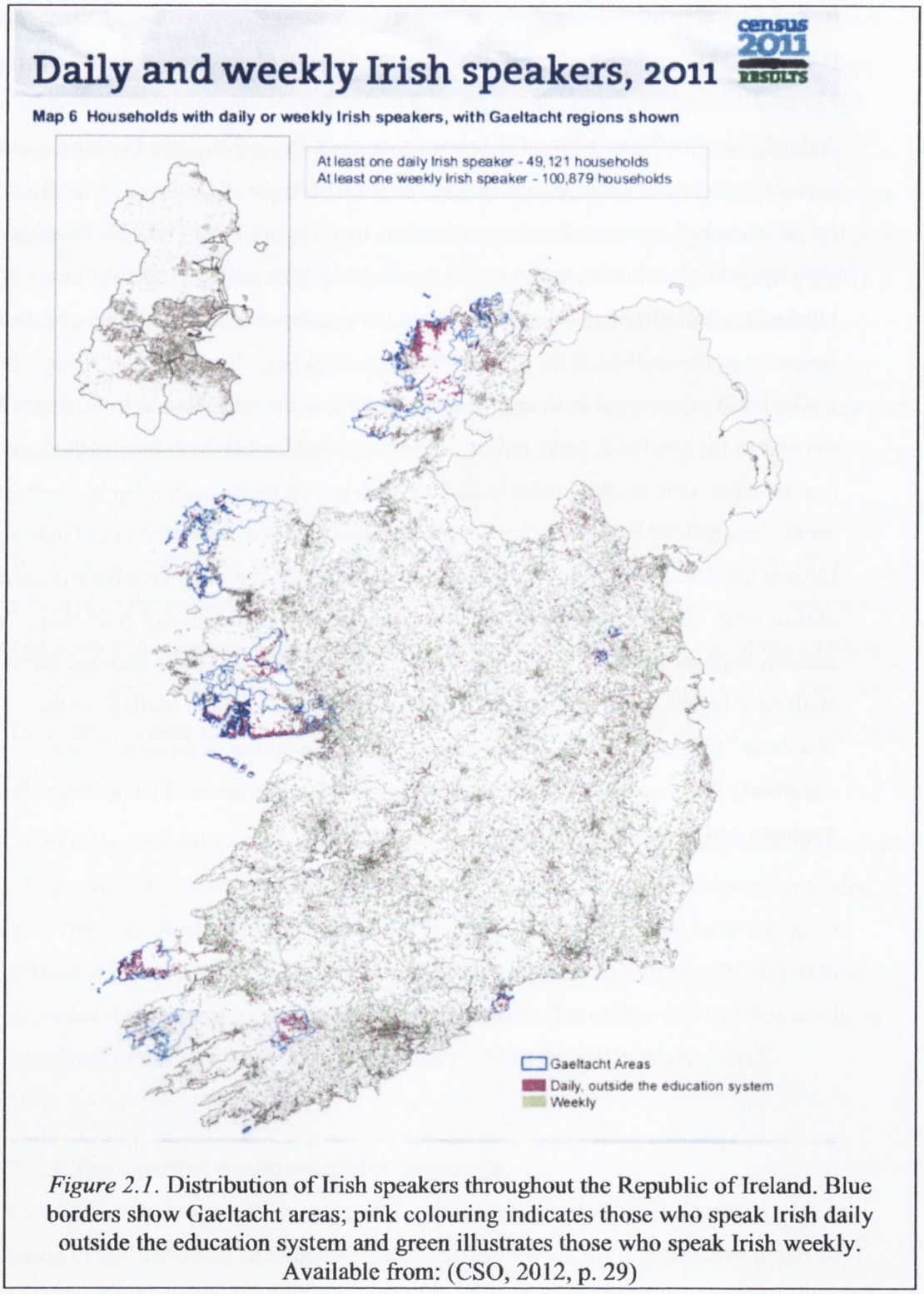


Figure 2.1. Distribution of Irish speakers throughout the Republic of Ireland. Blue borders show Gaeltacht areas; pink colouring indicates those who speak Irish daily outside the education system and green illustrates those who speak Irish weekly. Available from: (CSO, 2012, p. 29)

Repeated surveys since the 1970s have found a positive attitude to Irish among the population in general. The most recent Ipsos-MRBI (July 2010) survey conducted on behalf of Irish language organisations – Conradh na Gaeilge, Gael Linn, Comhar na Múinteoirí Gaeilge, Gaelscoileanna and Comhdháil Náisiúnta na Gaeilge – shows 57% of the participants thought it was important that children in Ireland should be brought up learning Irish. 19% were undecided and 23% thought it was not important. Mac Gréil and Rhatigan (2009) found that 41% believed that Irish should be revived for public use, 52% believed that the language should be preserved and 7% believe that it should be discarded. Of the 41% who believe Irish should be revived for public use, 4.7% believe in a bilingual system with Irish as the most widely used language, 32% believe in a bilingual system with English as the main language and 3.7% believe that Irish should be the main language of the country. This largely positive attitude to Irish is in line with the many surveys carried out since 1973 when research conducted by the Committee on Irish Language Attitudes Research (CILAR) gathered extensive data. Amongst its claims was that people saw Irish as having a symbolic role in ethnic identification and, as such, a cultural value. 56% of those surveyed at the time had either a "strongly favourable" or a "somewhat favourable" attitude to Irish. The results of subsequent studies summarised in Ó Fathaigh and Ní Bhraoin (1996) show a consistency in a positive attitude towards Irish by just over half the population. These surveys were mainly conducted by Irish Marketing Surveys in 1978, 1979, 1980, 1982, and 1985. Ó Riagáin and Ó Gliosáin (1994), however, note the large degree of pessimism for the future of the language amongst the population in general but they also draw attention to the fact that there is a strong disagreement with the statement that "Irish is a dead language". These seemingly contradictory positions illustrate the complexity of the position of the Irish language.

2.2.5 Irish in schools.

Unsurprisingly, the ambivalence towards Irish found in the population in general is reflected in the attitude and usage patterns found amongst the school going population. Murtagh (2003) summarises the situation in the following terms "The responses to attitude/motivation questionnaire items suggest that while most students were not

negative towards Irish neither were they willing to invest great personal effort in learning it" (p. 105).

Irish is a compulsory subject of study for all pupils attending second level schools in the Republic. Certain exemptions are allowed for pupils who were reared outside the State up to the age of eleven; those who spend at least three continuous years outside the State since first enrolling in school; those with a specific learning disability and those with serious intellectual impairment (Circular Letter M 10/94, Department of Education and Science). While the study of Irish is compulsory it is not compulsory to take the Leaving Certificate (LC) examination in the subject. The statistics for 2010 show 44,942 candidates took the LC Irish examination. The precise overall number taking the examination is not clear since a substantial number may be repeating single subjects or combinations of subjects. The corresponding numbers taking English and Mathematics are 51,499 and 52,280 respectively (State Examinations Commission, 2011, p. 7). This suggests that approximately 14% fewer candidates take Irish than take Mathematics. It is unclear, what, if any, course of study is taken by the approximately 7,500 students who do not sit the LC Irish examination.

There are approximately 45,000 pupils currently attending all-Irish schools (Gaelscoileanna), where they study all their subjects through the medium of Irish, between both the Republic and the North of Ireland outside of Gaeltacht areas (September 2013 figures). These attend at 180 primary schools and 41 post primary schools. This shows a very substantial increase in the number of pupils attending all-Irish schools during the past decade. In the year 1999/2000 there were 134 primary schools and 32 post primary schools between the Republic and the North of Ireland catering for 26,000 pupils (Gaelscoileanna, 2013). Figure 2.2 below shows the increase in the numbers of Gaelscoileanna in the 26 counties outside the Gaeltacht between 1972 and 2013 (Gaelscoileanna, 2013). There are a further 143 primary schools and 30 post primary schools in Gaeltacht areas (Mac Donnacha, 2005).

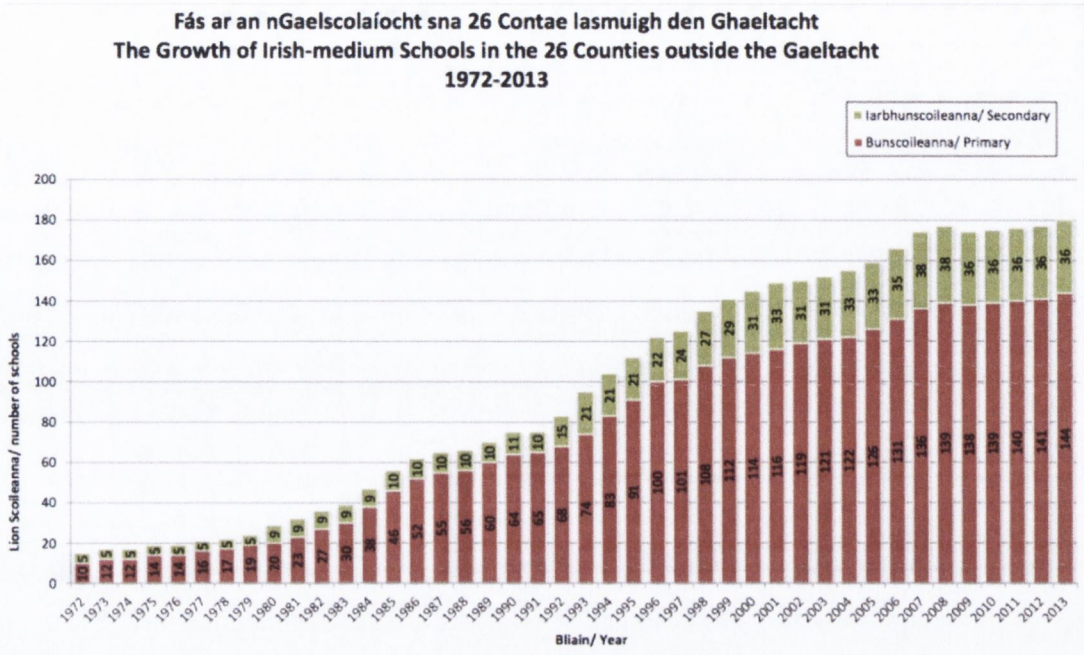


Figure 2.2. Number of Gaelscoileanna in the 26 Counties outside the Gaeltacht 1972-2013. This figure illustrates the growth in the number of Irish-medium schools in the Republic of Ireland (outside of Gaeltacht areas) – from 10 primary and 5 second level schools in 1972 to 144 primary and 36 second level schools in 2013. Available online from: http://www.gaelscoileanna.ie/files/Growth.F%C3%A1s-gscoil-72-13.pdf

2.2.6 The Irish language syllabus for second level schools.

The distinction between Gaelscoileanna, which are located outside Gaeltacht areas, and Scoileanna Gaeltachta, located in the Gaeltacht, may be more apparent than real. Mac Donnacha, Ní Chualáin, Ní Shéaghdha and Ní Mhainín (2005) found that 1/4 of all children in Gaeltacht areas were born and spent their early years outside the area. They also found that 46% of school-going children in core Gaeltacht areas start school with little or no Irish. Harris, Forde, Archer, Nic Fheararaile and O'Gorman (2006) found that children from Gaelscoileanna outperformed their Gaeltacht counterparts in most language ability tests and that there was a significant decline in the standard of Irish amongst children in Gaeltacht schools over the period 1985 – 2002. This was similar to the decline in the standard of Irish in English-medium schools over the same period. The same

decline did not hold for Gaelscoileanna even though their numbers increased from 1.1% of total schools in 1985 to over 5% by 2002.

There has been a general belief that if the appropriate teaching methodology could be discovered/devised then the Irish language could be acquired by almost all of the school going population. Teaching methodologies have followed the prevailing views of the time from direct methods through audio-visual methods to communicative approaches. Language teaching and literature teaching have been intertwined in a way which is unusual for second/foreign language teaching syllabi. No distinction is made between courses for those who have been reared entirely through Irish, those who came through all-Irish schooling and those for whom it is merely a daily slot on their school timetable. The failure to allow for syllabus variation would seem most likely connected to the over-arching importance attached to the Leaving Certificate as the competitive examination which determines entry to the more and less prestigious faculties in Universities. Those who have reared their children through Irish are unwilling to surrender their competitive advantage in the Leaving Certificate. The stranglehold of the examination prevents the possibility of variation in syllabi since the variants may not be seen to be of equal difficulty and hence a source of inequality among the school going population.

There is a huge variation in the contexts in which Irish is taught at second level schools throughout the country. Amongst these contexts are:

(a) huge variation in language backgrounds amongst the pupils ranging from traditional native speakers to those who have little more than a few words

(b) variation in how pupils see the language going from 'of no use' to a very significant aspect of national identity

(c) variation in the communicative function of Irish outside the school. Unlike the Welsh situation there are no shops, restaurants, or other social outlets where Irish

is the sole acceptable language. Added to this is the fact that there are no monolingual speakers of Irish left

(d) there is huge variation in the levels of proficiency amongst teachers ranging from relatively low communicative competence to traditional native speakers

Modern language teaching theory claims that languages are best acquired when the learners are immersed in situations where the target language needs to be used to achieve some genuine communicative purpose. Given the absence of a truly monolingual Irish language context, the teaching of Irish to suit non-existent situations outside of school seems somewhat unrealistic. In order to teach a language which has a functional communicative purpose, the creation of an attractive virtual world seems appropriate. In this virtual world there is need for providing high quality native-like voices where the learners can become active participants in the situations that are created. The present thesis is concerned with exploring the attractiveness and value of serious games as an aid to teaching Irish in second level schools in Ireland.

2.2.7 The use of technology in Irish language teaching/learning.

To date there has been relatively little interactive digital-based material available for the teaching of Irish in second level schools. Fios Feasa has produced interactive learning materials for the development of reading skills, mainly for primary school children (www.fiosfeasa.com).

There is some other teaching material available on CD-Rom such as Turas Teanga, where material is presented by way of cartoons and photographs in a realistic setting with high quality film production and accompanying explanatory written material and exercises. This is directed primarily at adult beginner learners.

Dréimire (The Council for the Curriculum, Examinations and Assessment, Northern Ireland) has produced a series of restricted interactive materials, with voice optional, for teaching initial reading to primary school children.

An exhaustive list of teaching materials is published by Gaelscoileanna and is available at http://www.gaelscoileanna.ie/resources/. While digital-based teaching materials are becoming more available as technology becomes more integrated into the school system, one cannot say that there are any genuinely interactive materials available for the teaching of Irish.

2.2.8 Irish in the context of CALL for endangered languages.

Ward (2001), who proposed a template for the development of a CALL program for Endangered Languages, referred to a lack of CALL materials for such languages at that time. One cannot claim that there have been any significant developments in this regard in the intervening years. Some attention has been paid in the literature to the function of CALL in less widely used languages. Ward (2004) lists four benefits which CALL can bring to such situations. She claims that:

(a) CALL can change negative attitudes towards the language.

It has often been pointed out that learners of Irish in schools have been put off by antiquated teaching methods and teaching materials, with an emphasis on the literature of times gone by in remote parts of the west of Ireland. The types of platforms developed for this study are entirely different in nature and depart from these older models. They effectively aim to be more in keeping with the technology-rich environment in which younger learners now live. To this extent, we can argue with Ward that these CALL approaches, particularly game-based interactive activities with good graphics, are likely to have a positive impact on attitudes towards the language among younger learners.

(b) It can help arouse interest in the language.

This relates to the above point insofar as the activities involved in these types of CALL platforms are likely to engage the attention of younger learners.

(c) It can help document languages which are in danger of extinction.

This is a point that has been made repeatedly by the *ABAIR* research group in relation to the broader provision of speech and language technology for the minority and endangered languages. In providing access to synthetic speakers of the dialects of Irish, learners from any part of the globe have access to the spoken language.

(d) It may lead to language maintenance.

This is a corollary of the previous point. A speech synthesis system provides and preserves a permanent record of the language as it is spoken today. The *ABAIR* group aims to provide synthetic voices even for dialects for which relatively few speakers remain. In the future, even if no native speakers remain, the synthetic version provides a resource, which will remain there for those who wish to study or acquire that dialect. CALL applications which use synthetic voices contribute in a similar way.

Advantages pointed out by other researchers in the field are:

(e) It can make endangered languages more readily available to younger learners (Auld, 2002).

This is undoubtedly the case. Availability of speech technology and of CALL applications on the web makes access very easy, not only for younger learners but also for a wider learning public. The experience of the *ABAIR* group has been that the voices available on the web are accessed from around the world.

(f) It can help to "glamorise" the language and give it a parity with the more glamorous media products which are available in the major world languages (Edwards, Pemberton, Knight, & Monaghan, 2002).

This point relates to the discussion above concerning these kinds of CALL platforms being more in tune with the activities of today's young generation. Multimedia games and activities are widely available in the major languages, and it is crucial for an endangered language to be able to match the attractive offerings of other languages.

(g) It can facilitate indigenous language stabilisation and development (Jancewicz & MacKenzie, 2002).

This is partly the same point as was made in (d) above. It goes without saying that stabilisation and development of an endangered language depend crucially on how successfully it can be taught. For all the reasons mentioned, CALL applications have a potentially powerful role to play in this process.

Although Irish is an endangered language, and under-resourced in terms of the development of speech and language technology and CALL applications, it enjoys Constitutional protection from the State. It is one hope of the present study, as well as of the wider *ABAIR* project, that what is being done here to develop resources and applications for Irish, can serve as a model for other endangered languages.

2.2.9 Summary.

The position of the Irish language is complex. Although it has enjoyed very considerable Constitutional and official institutional support, the population in general has a somewhat ambivalent attitude towards it. While there is support for the language, and while over 40% of the population claims to have Irish, this support does not extend to a widespread daily use of the language, outside of the education system. There are no monolingual speakers of Irish surviving.

A similar ambivalence to that of the general population exists in schools. More than 14% of the cohort does not take Irish in the Leaving Certificate even though it is intended to be a compulsory subject. There is, however, a very significant growth in the Gaelscoileanna sector where all subjects are taught through Irish.

Since there is no area of life where Irish is the only available language, smart digitalised environments offer learners the possibility of using the language in a quasi-realistic setting and may increase their motivation as well as giving a more youthful and glamourous image to it. There are few interactive digital Irish language facilities available and the present study aims to develop prototypes of such facilities. It is intended that such prototypes be highly interactional in the sense that the learner can act and react towards the virtual world situations giving at least a semblance of genuine communicative experiences. The following sections delineate the theories of language acquisition and gaming that underpin these prototypes. Section 2.3 of this thesis explores the raison d'être for interactive task-based activities as an aid to second language learning. It focuses specifically on how CALL can be used as a facilitator of such activity.

2.3 CALL and the Interactionist model of SLA

2.3.1 Overview.

This section explores the concept of CALL as it has evolved over its relatively brief history. It looks at the congruence between technological development and the ever-increasing ability of the computer to handle material of a more complex nature. It examines how CALL has responded to various theories of second language teaching/learning which became prevalent over the years. It looks in particular at the CALL response to the behaviourist theories of language acquisition which underpinned the audio-visual teaching methodology. It further explores how CALL was able to respond to the higher demands of communicative language teaching methodologies and

in particular to CALL as a facilitator which allows the learner to interact with the computer in novel ways. Such activities are frequently referred to as belonging to an interactionist model of second language acquisition which is given expression by communicative approaches to language teaching/learning.

2.3.2 Computer-Assisted Language Learning.

CALL is not yet a clearly defined area of study. It is still at a developmental stage where there is debate among scholars as to its precise range and scope. Definitions of CALL vary as indeed does the terminology associated with the use of computers as an aid to language learning, CALI (*Computer-Assisted Language Instruction*), CAI (*Computer-Aided Instruction*), TELL (*Technology-Enhanced Language Learning*) etc. (Beatty, 2013). Levy and Hubbard (2005) have a more exhaustive list of CALL-related terms which have gained favour at various times but a detailed consideration of them is unnecessary in the present context.

I do not propose here to delineate a detailed history of CALL since this has been researched exhaustively by numerous students of CALL (Bax, 2003; Davies, Walker, Rendall, & Hewer, 2012; Davies, 2006; Liu, Moore, Graham, & Lee, 2003; Warschauer & Healey, 1998). This intense interest in the history of CALL reflects an unease among researchers about the various definitions of CALL and the lack of a clear vision of its future direction.

At its most generalised level CALL can refer to any use of a computer as an aid to language learning. Levy's (1997) definition went largely unchallenged for a number of years having been accepted by professional associations such as Eurocall, CALICO, etc. His definition saw CALL as "the search for and study of applications of the computer in language teaching and learning" (Levy, 1997, p. 1). Programs were designed by "letting computer whizzes first explore how the computer can do something particularly well and then design the program to take advantage of that" (Underwood, 1984, p. 83). This may

be seen as putting the cart before the horse since its emphasis is on new ways to use the computer rather than on the language learning process and the degree to which computers may be of assistance to this process. Levy (1997) makes the distinction between CALL as a tutor (i.e. having an interactive dimension) and CALL as a tool (efficient provider of backups such as dictionaries, spellchecks, etc.). CALL as tutor was seen by Levy as a temporary substitute for a human teacher.

The CALL programs which existed in the 1990s and earlier, were relatively rigid and determined the course of a lesson in a predefined way. In more recent years, however, greater interactivity between learner and computer has become possible, thereby reducing the controlling role of the computer program. CALL ('tutorial' or 'tool') may be seen as an aid to language learning, rather than a replacement for a teacher, since education is essentially an interpersonal process where the teacher cannot be substituted for in any meaningful way (Yugandhar, Srinivas, & Ramakrishna Rao, 2010). In this scenario, Levy's distinction between tutor and tool functions becomes unhelpful as both can be seen as tools or aids to the language learning process.

In this thesis, I argue for the need for developments on the technological side to come as a response to pedagogical innovation and for collaboration between the pedagogues and engineers of both the software and hardware varieties. Colpaert (2006) regretted the lack of coordination between pedagogues and engineers. A more cooperative spirit appears to be developing between them in the very recent past, however.

Clark and Mayer (2008) see e-learning in general as "training delivered on a computer (including CD Rom, Internet, or intranet) that is designed to support individual learning or organizational performance goals" (p. 7). No distinction is made here between the computer's function as a 'tutor' or a 'tool'. This puts the emphasis back on the learning, or the cognitive modification experienced by the learners, as the central focus is on the endeavour, with the computer acting as an aid rather than an end. Chapelle and

Jamieson's (2008) definition of CALL as "the area of applied linguistics concerned with the use of computers for teaching and learning a second language" (p. 1) brings CALL into line with Clark and Mayer's definition of generalised e-learning, allowing for computers to be seen as an aid to the language teaching/learning process. I am taking it that the term 'second language' here refers to second or subsequent languages.

The term *Tutorial CALL,* which refers to any language teaching programme which is pedagogically driven and has an appropriate specific technological underpinning, developed somewhat of a bad reputation among teachers for a variety of reasons (Hubbard & Bradin Siskin, 2004, p. 454). It was primarily useful for language drills (drill and kill), and it did not live up to the high expectations which teachers had for it. "It is dominated by commercial interests" and "the teacher's role seems limited" in the language teaching/learning process, are among the criticisms proffered. This study aims to develop material which surmounts these weaknesses. In the following sections the relationship between CALL programs and prevailing theories of pedagogy and technological developments is noted.

The present approach puts the emphasis on the learner acquiring a language (no distinction is being made at present between language acquisition and language learning as postulated by (Krashen, 1981)) and a computer being used to facilitate this process. There are three key considerations which must define approaches to CALL development, namely:

(a) the prevailing theory or theories of language acquisition

(b) the current state of technological development, i.e. what is within the capability of the computer

(c) learners' reactions to the use of computers as a learning aid, i.e. the engagement of learners with the computer at an affective level or indeed its influence in their general motivation to acquire the target language (Dickinson, 1995)

2.3.3 Language Acquisition Theories and CALL.

2.3.3.1 Early approaches to CALL.

By coincidence CALL programs have fitted reasonably well with prevailing theories of language acquisition over the years. Audiolingual/audiovisual approaches to language teaching relied on theories of behaviourism for their intellectual underpinning. Language was regarded as observable behaviour which could be studied without reference to hypothetical mental activity (Sampson, 1980). Language teaching placed an emphasis on oral/aural skills and used aids "...including mechanical apparatus, such as motion pictures, phonographic records, recording machines, magnetic tape recorders, radio and telephone" (Agard et al., 1945, p. 156).

A behaviourist model of language acquisition was to be found either explicitly or implicitly underlying a great many of the approaches to language teaching up to the 1970s. As early as 1926 Sydney L. Pressey developed a system of rotating cylindrical drums whereby students could respond to multiple-choice questions by pressing one of four keys, each of which represented one of the four possible answers (Pressey, 1926). This has been seen as a precursor to CALL programs (Beatty, 2003).

PLATO (Programmed Logic for Automated Teaching Operations), originally developed in the University of Illinois in the early 1960s, represented the first major technological development which made CALL possible. The program had three phases of development between 1963 and 1969 and introduced a programming language (TUTOR), which allowed non-computer expert teachers to design their own lessons on computer. The programs developed were of a 'drill' nature but many could give feedback to students to help them arrive at the correct solution (Salaberry, 2001). The program represents one of the first forms of interactive computing and was the first attempt at Natural Language Processing.

Other significant early CALL programs include CLEF, a French grammar revision package developed in the late 1970s by a consortium of Canadian universities. Published by Camsoft, it claimed to give the learner "clever feedback routines" and added colour and animation to emphasise important features. It is based on limited routines and if the learner does not succeed in getting the correct answer after the second attempt, a consoling message is given (CLEF, 1997). "Fun with Texts", first published by Camsoft in 1985, is still widely in use in British schools and its fourth edition remains in demand to the present. It allows the teacher to download texts from external sources and then creates text manipulation exercises such as text scrambling, cloze tests and decoding activities for the students (Davies & Bruzzone, 2004). This type of program, though close to the typical textbook exercises, extends the variety of texts which can be made available for purposeful study by the students. The use of authentic text is favoured by many language teachers (Breen & Candlin, 1987; Devitt, 1997; Little, Devitt, & Singleton, 1989; Mishan, 2005). One cannot claim that these programs have a distinctive philosophical underpinning, in the sense that they are based on distinctive theories of language, or of language acquisition.

2.3.3.2 Communicative approaches to language acquisition and CALL.

Interactionist theories of language acquisition claim that language is acquired through meaningful interactions with others in the target language. Interactionist theories underlie what is referred to as 'communicative approaches to language learning/teaching'. These approaches are generally recognised to have gained prominence during the 1970s (Littlewood, 1981; Richards, 2006). While the roots of these approaches were diverse and interesting, I do not propose to delineate them in the present review. Underpinning communicative approaches is the idea that language is a system of communication (Richards & Rodgers, 2001). It is claimed that language should be taught using realistic settings and contexts. The aim of language learning is to achieve language competence, where competence is defined as "the expression, interpretation and negotiation of meaning" (Savignon, 2006, p. 673). Learners are guided through situations which are

intended to be meaningful and realistic to them and they are expected to participate appropriately in linguistic encounters.

Interactionist Model of SLA and its Assumptions.

Pica (1997) investigates relationships between SLA research and practice. She categorises various strands of research and concludes that the type of SLA research most significant for CALL design is that which *complements* L2 instruction. Chapelle (1998) builds on Pica's (1997) paper and discusses the interactionist model of SLA in terms of a theory of SLA drawing together hypotheses that are implied in this theory, including 'input saliency, opportunities for interaction, and learner focus on communication'. A model is presented as part of her paper which shows the basic components in the SLA process, summarising a consensus view among interactionist SLA researchers. In this model it is hypothesised that target language input acts as the potential starting point for acquiring aspects of the L2. The model is reproduced below in Figure 2.3 and an explanation of the components of the model follows:

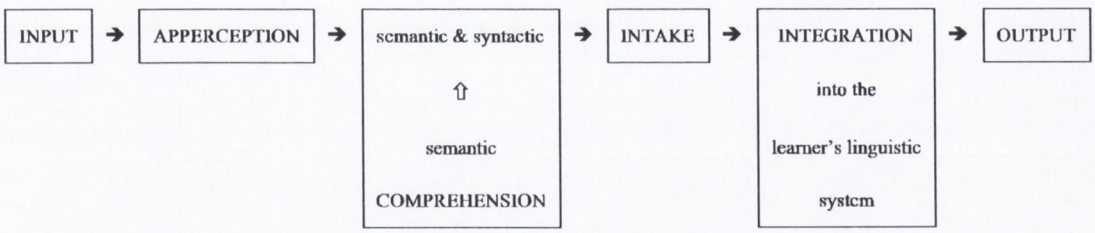


Figure 2.3: Basic components in the SLA process in interactionist research (taken from Chapelle 1998)

The INPUT stage of this process refers to the target language to which the learner is exposed. Chapelle's model expands considerably on Krashen's (1982) idea that a lot of comprehensible input is what is needed for SLA insofar as it is accepted that much of the target language input goes over the language learners' head and that *'noticing'* aspects of the target input (the APPERCEPTION stage) is a crucial component.

COMPREHENSION deals with the hypothesis that learners' acquisition of the syntactic system is not necessarily going to be helped by semantic comprehension alone – recognition of isolated lexical items and interpretation of non-linguistic cues can at times be sufficient to understand a message. This leads to the next stage of the model where comprehension takes place through a combination of semantic and syntactic processing and the input can become INTAKE – i.e. language that has the potential to develop the learners' linguistic system.

Chapelle describes INTEGRATION as processes for using the 'intake' (that's in learners' short term memory) to influence the development of the linguistic system, which in turn affects the learners' L2 output.

OUTPUT in itself is an important stage in learners' linguistic development as it forces the learner to use the syntactic system and therefore to develop this aspect of their L2. Output also naturally elicits further input from interlocutors. If there is a problem with the learner's output an indication of this might be given by the interlocutor which will result in the learner '*noticing*' aspects of linguistic form and going through the stages of the model again in order to produce more output. This process is referred to as *negotiation of meaning* (Long, 1996) and is believed to facilitate L2 development.

When designing instructional materials, the importance of including features that prompt learners to *notice* important aspects of the L2, is widely accepted. Many different approaches have been taken in the field of CALL in recent years in an attempt to achieve this goal. The emergence of various new technologies enabled researchers to investigate how new materials could best be exploited for language learning purposes while maintaining the inextricable link with the relevant language learning theories. The emergence of Digital Video Discs (DVDs) offered greater access to video clips and allowed various combinations of frames to be accessed very quickly. Greater interaction

between user and computer became possible and students could come to control the direction of the lesson. Amongst the more significant programs to emerge in the 1980s were: *Macario; Montevidisco; Interactive Digame;* and the *Athena Language Learning Project* (Hubbard, 2009). Their effectiveness has been the subject of some debate (Liddell & Garrett, 2004). Bax (2003) did not find the programs to be communicative in any significant way (p. 18) because they lacked the central features of human communication and interaction.

If *interpersonal communicative situations* was the demand then the development of the internet offered a technical solution to that demand. In his extensive review of CALL literature, Jung (2005) traces the first references to the internet and the World Wide Web to 1993. He sees this as the watershed between online and offline media. In the years subsequent to this, online approaches to CALL enjoyed an exponential growth (Otto & Hatasa, 2006). Felix (2003) examines the usefulness of the web for distance learning. The fact that she conducted a further study in 2004 into the disposition of students to web-based teaching/learning, is indicative of the 'newness' of this type of technology and the uncertainty about its future. Teachers were able to set up cross-cultural projects for their students and language learning ceased to be an exclusively classroom based activity. Authentic materials now became available and the ideals of a communicative approach to language acquisition became achievable.

A survey of the CALL literature carried out for the present study shows that since the year 2000 researchers' attention has been predominantly directed to online language learning projects where each emerging technology is exploited for its potential to facilitate language learning. González-Bueno (1998) reviewed the effectiveness of using email as a tool to promote foreign language learning in and out of the classroom. Osuna and Meskill (1998) found the web to be a suitable tool to increase language and cultural knowledge as well as to increase motivation. New terminology emerged such as computer-mediated communication (CMC) and questions were posed as to the relative merits of synchronous (real time) or asynchronous (deferred time) communication

(Blake, 2007). More recent literature includes studies of: electronic conferencing (Fitze, 2006); blogs (Miceli, Murray, & Kennedy, 2010); mobile blogs for the exchange of files (audio, video, images and short texts) between mobile phones (Comas-Quinn, Mardomingo, & Valentine, 2009); blogs for specific purposes (Murray & Hourigan, 2008), i.e. a specific social networking tool, (Hourigan & Murray, 2010); podcasting (Ducate & Lomicka, 2009); various mobile devices which facilitate learners' reading comprehension (Chang & Hsu, 2011); Web 2.0 technologies, where new content can be created in the "writable web" (Kárpáti, 2009) and the use of the Microsoft Speech Application Software Development Kit (SASDK) for the development of oral skills (Chen, 2011).

Most of the above computer-based resources are based on setting up communication systems between competent speakers of the target language and learners. They facilitate real communication where the learner needs to negotiate meaning with the competent speaker. As demonstrated above with Chapelle's (1998) model, interactionist theories suggest that this is the most effective method of language acquisition. In the case of endangered languages, such as Irish, where there are few, if any, native monolingual speakers left, such resources are of relatively little value because the native speakers would have little or nothing to gain from the interactions and they would be very unlikely to persist. In this context, the present project aims to develop prototype interactive materials where the computer can take the place of the expert native speaker and can react with the learner in human-like formats where speech turns can be relatively unique and unpredictable. Such exercises are being referred to as 'tasks' and the learner is given specific 'tasks' to complete, be this in a game-like scenario, or in some kind of dialogue-based interaction. In attempting to develop such materials one is placing the project within the realm of 'task-based language learning', which assumes that the language is acquired most effectively when it is used communicatively in order to achieve some given task.

This brief review of the literature gives an indication of the trends in CALL research as opposed to being an exhaustive bibliographic statistical review. Some interesting trends emerge. Researchers are now more concerned with the effectiveness of the technologies as tools for language teaching/language acquisition rather than with the technologies involved in the creation of the various teaching/learning platforms. CALL is being taken back from the engineers by the pedagogues. Engineers and pedagogues are adapting a more cooperative or interdisciplinary approach in the development of Tutorial CALL programs.

2.3.4 Summary.

This section has reviewed the emergence of CALL and its connection to the prevailing language acquisition theories. In brief, the behaviourist theories of language acquisition, which were commonly accepted pre-1970s, required a relatively unsophisticated computer program to underpin them. These were primarily based on stimulus and response. The computer offered great possibilities when it came to 'drill' and could outdo the patience of any teacher. As 'interactionist' or 'communicative' approaches to language acquisition became prominent, more sophisticated interactions became technically possible and the internet offered the possibility of matching learners with native speakers. Genuine communication and negotiation of meaning became possible. Where there are no monolingual speakers of the target language, such communication systems are not practical. For this reason one must create platforms where the computer can become a native-like interlocutor with the learner in a virtual world. The learner sets out to achieve some predetermined task, which requires interacting in novel ways with the computer. Such activities are referred to as 'task-based language learning', which comes under the general umbrella of 'communicative language teaching/learning'.

2.4 Task-based Language Learning and Teaching

2.4.1 Overview.

This section reviews the key features of task-based language learning (TBLL) in order to set the parameters which underpin the three platforms being developed for the present project. Some definitions of TBLL are explored and the connection between these and computer gaming as a legitimate CALL activity becomes more apparent.

2.4.2 Recent developments in communicative language teaching and CALL.

Communicative approaches to language teaching have remained dominant to the present day with a number of specific concepts coming to the fore (Brown, 2007). Most prominent amongst these is *Task-based Language Learning* where 'task' refers to "a contextualised, standardised activity which requires learners to use language, with emphasis on meaning, and with a connection to the real world, to attain an objective" (Bygate, Skehan, & Swain, 2001, p. 12).

Klapper (2003) sees task-based language teaching as a further development of communicative language teaching (CLT). He maintains that the communicative interaction characteristic of task-based work "*provides sufficient comprehensible input to 'trigger' acquisitional processes.*" This follows from the Chapelle (1997) model, demonstrated in Figure 2.3 above, which illustrates that language learning depends on learners being involved in real communication in which they use language in a meaningful way. Klapper (2003) defines tasks as "meaning-based activities closely related to learners' actual communicative needs and with some real-world relationship, in which learners have to achieve a genuine outcome (solve a problem, reach a consensus, complete a puzzle, *play a game*, etc.) and in which effective completion of the task is accorded priority" (p.35).

Prabhu (1987) referred to the cognitive processes pertaining to "a task", in that it should involve the learner in making connections between pieces of information, deducing new information and evaluating it. Ellis (2003) sees tasks as involving selecting, reasoning, classifying, sequencing and transforming information from one form of representation to another. TBLL ideals led to the development of various games for the classroom involving pair-work and group activities where learners were encouraged to solve problems presented in the target language, either individually or as teams (Ellis, 2001; Julkunen, 2001).

Doughty and Long (2003) identify ten task-based language teaching methodological principles which have applicability to CALL. These aim to serve as a basis for creating the optimal psycholinguistic environments for computer-based language teaching. Many of these principles had long been accepted as a general theoretical underpinning for second language teaching. These include building lessons around tasks rather than texts, promoting learning by doing, providing appropriate feedback, and showing respect for mental stage of development. Their most significant contribution to the debate is their attempt to operationalise these principles into design features of CALL programs. They suggest that ultimately the developmental sequencing in language teaching could be arranged according to a gradation of task complexity. Such an approach, while intuitively attractive, has not been widely used in practice.

Gaming as a CALL activity can be seen as being a TBLL opportunity where the learner needs to use the target language appropriately in order to navigate through the game. In this way, as discussed above in Section 2.3.3.2, it provides the conditions which the interactionist theory suggests are required for language learning to take place. Consistent with this, Gee (2005b) sees the potential for video games to enhance general learning skills. The problem solving methodologies encountered in a game have a general application to real-life situations. The player forms goals and in order to reach these goals he/she must recognise problems and solve them within the constraints of the world

created in the game. The player must figure out the rule system in order to solve the problems and achieve the goals.

Gee's (2005b) description of the learning process associated with computer gaming mirrors that of Ellis's (2003) framework for describing the essentials of TBLL. A goal is set for the task; input of verbal and non-verbal information is supplied; conditions are set down as to how the information is used; the methodological procedures to be followed in performing the task are set down and outcomes are evaluated in terms of the extent to which the goals are achieved. Mitchell and Savill-Smith (2004) summarise the conclusions of earlier studies on the effectiveness of computer games to promote learning. They see their advantages in terms of their ability to motivate learners; their encouragement of active learning and their ability to give instant feedback to what has been dubbed "screenagers" (a word which entered the Oxford English Dictionary's 2003 edition). Tong and Cadwell-Harris (2008) note that the effectiveness of using virtual interactive objects by adult language learners remains largely unexplored and speculate that a virtual environment with humanoid objects that respond to user commands would be highly interactive and possibly highly rewarding. This contention can trace its philosophical underpinning to Dewey's (1910) assertion that the purpose of language is to coordinate action. The functional psychology espoused by Dewey sees cognitive processing in terms of active adaptation to the person's environment. Zheng, Young, Wagner and Brewer (2009), in following in the functional psychology tradition, see language as a sociocultural behaviour and the construction of new knowledge as both a cognitive and social process. Interaction in virtual environments, it is claimed, may have the same legitimacy for the player/participant as that of 'real world' encounters.

2.4.3 Summary.

This section has reviewed the definitions of 'tasks' in the context of interactionist language learning. Essentially tasks involve the learner using existing information in order to deduce new information, which in turn leads to the achievement of some goal. Task, as viewed here, is very close to involving the same essential processes as those

involved in computer gaming. It is postulated in the present thesis that computer games can be an effective means of creating the conditions where task-based language learning activities can be created. A closer examination of game-based language learning and the progress made to date in this domain follows in Section 2.5.

2.5 Game-based Language Learning

2.5.1 Overview.

Task-based language learning is at present the dominant language learning theory and the one being subscribed to in the present thesis. A detailed consideration of the concepts involved was undertaken in Section 2.4. The present section aims to operationalise the concept of task-based learning in the context of virtual reality environments, including computer games (see Section 2.5.2), virtual worlds (see Section 2.5.3) and chatbots (see Section 2.5.4).

The concept of 'gaming' in general is examined and games which have a purely entertainment value are contrasted with those which have educational or informational intent. The notion of what is now referred to as 'virtual worlds' is explored. This section also introduces the concept of 'chatbots' in a game-like scenario and their possible contribution to CALL. The present thesis makes the postulation that all three can be integrated in order to produce effective CALL game-like activities.

Features which make general computer games effective are identified (see Section 2.5.5). This section goes from the design features of games in general (see Section 2.5.6) to identifying those features which make for effective CALL games (see Section 2.5.7). A summary is included in Section 2.5.8.

2.5.2. Background: computer games and learning.

The term 'computer game' is being used here in its broadest sense, in referring to any computer-based activity which has an interactive dimension, where the game player is actively involved in achieving some predetermined aim.

Gee (2005b) makes a type of banner headline opening to his essay, "Video Games are Good for Learning". He quickly modifies the banner claim by admitting that different types of games can have different effects. The danger of overgeneralising the value of games is flagged by Van Eck (2006, p. 17):

> *If we continue to preach only that games can be effective, we run the risk of creating the impression that all games are good for all learners and for all learning outcomes, which is categorically not the case.*

More than anyone else Gee, in a series of papers and books, has developed the proposition that computer games replicate the way the human mind works (Gee, 1992, 2005a, 2005c, 2007, 2008a, 2008b). This view of the human mind is based on the concepts of connectionism as established in general cognitive science. Connectionism tries to explain human intellectual abilities using artificial neural networks. The brain has large numbers of connections with different weights that measure the strength of the connections between the units (Garson, 2012). Computer games are similarly composed of sets of weighted interconnections which simulate the functioning of the human brain. There is an assumption that the computer game can emulate the brain and therefore optimise the conditions of learning which the brain requires.

Few, if any, researchers make the strong assertion that all computer games are good for learning. Instead, the focus of attention is on which elements of video games make them attractive and engaging and on how these elements may be exploited for educational

purposes (Moreno-Ger, Burgos, Martínez-Ortiz, Sierra, & Fernández-Manjón, 2008). The predominant strength of the gaming approach to teaching/learning is its ability to motivate the learner and produce a stimulating, fun environment while developing the desired knowledge and skills (Amory, Naicker, Vincent, & Adams, 1999; Kirriemuir & McFarland, 2004; McCosker, 2009; A. Mitchell & Savill-Smith, 2004). The intensity with which the learner/computer gamer engages with the computer is referred to as "flow" (Csíkszentmihályi, 1988). It was postulated that:

> *When a person's skill is just right to cope with the demands of a situation – and when compared to the entirety of everyday life the demands are above average – the quality of the experience improves noticeably. It does not matter whether one originally wanted to do the activity, whether one expected to enjoy it or not. Even a frustrating job may suddenly become exciting if one hits upon the right balance* (Csíkszentmihályi, 1988, p. 32).

This is a more modest proposal in relation to the concept of flow than that made by subsequent analysts who attempted to link flow to computer games. Franciosi (2011) characterises it in terms of a loss of self-awareness and a distortion of time. He goes on to claim that flow theory can also be applied to language learning since it too can be considered a challenging endeavour. There is a very great difference between a war game where the player in the persona of one of the game characters is at risk of being blown up and killed because of one false move, on the one hand, and language acquisition on the other. The immediacy of the former can lead to flows of adrenalin but one also knows that "no one ever died of a split infinitive" Silberman (1970, p. 436) and so the two activities are qualitatively different and cannot be directly compared to each other.

2.5.2.1 Entertainment or 'serious' games.

Literature on computer games distinguishes between those games which are intended for entertainment purposes and those with some type of educational purpose. Digital games used for purposes other than mere entertainment have been dubbed "serious games" (Susi, Johannesson, & Backlund, 2007) or "edutainment".

Ratan and Ritterfeld (2009), in a survey of serious games from 1997 to 2007, defined 48% of games as having a primary learning principle of practical skills and they tended to impart these skills by way of repetition and practice. 24% of games had cognitive problem solving as their learning principle. In terms of content, 9% were associated with occupational related matters, such as the knowledge and skills associated with specific occupations. 8% were directed at health related content so as to impart knowledge and habits associated with health improvement. 5% of games related to military knowledge and skills and games with marketing related content, such as brand awareness, accounted for less than 1% of the market. Games associated with CALL did not feature, a fact which shows their miniscule representation in the market and that their development as task-based language learning aids is largely at the prototype stage and the systematic design and implementation of these games still remains an elusive topic (Moreno-Ger et al., 2008). The present project aims to identify some of the elements of computer-based language teaching activities which are particularly useful to a language learner. Since the development of games as a CALL tool is still in its infancy there are, as yet, no comprehensive criteria which are generally acceptable.

In order to examine the criteria most likely to lead to an effective language learning game I am making the assumption that task-based language learning, as the present prevailing theory, is the activity which is being promoted (see Section 2.4). Effective games will therefore share some of the characteristics which are common to all effective computer games and some which are unique to language learning. In fact, there need not

necessarily be a clear distinction between 'serious' and 'non-serious' games. Many of the same characteristics are shared by both genres. Each provides some kind of learning experience for the player. Prensky (2007) goes as far as to make the strong assertion that "Anyone who makes a distinction between games and education clearly does not know the first thing about either of them" (Breuer & Bente, 2010, p. 90). In serious games there is an assumption that the gaming element is the central focus and that the activity itself is intrinsically motivating because it "provides fun" (Ritterfeld, Cody, & Vorderer, 2009, p. 4). The source of the enjoyment can vary depending on the user and the situation.

Effective general computer games share the characteristics of establishing clear goals for the game, rules for playing the game, task feasibility, self-governance and immediate feedback (Peirce & Wade, 2010). Games have a storyline which set a rationale for playing the game; game mechanics which determine how the world of the game operates; an immersive graphical environment, i.e. 2D/3D graphics, sound and animation; interactivity or the influence which the player's actions can have and risks and consequences (Derryberry, 2007).

2.5.3 SLA in virtual worlds.

Virtual world platforms have become increasingly important for the future development of CALL. It was claimed by Joe Miller, Vice President of Platform and Technology Development for the largest virtual world platform (the Linden Lab) that language learning is the most common education-based activity in Second Life (Team Engage, 2009). Most virtual platforms cannot be described as games since they do not have designated goals or objectives, rules of play, or criteria for winning. Virtual life platforms are continually evolving through the interactions of the participants, who have life-like social interactions while engaging in meaningful learning activities (Cooke-Plagwitz, 2008). The largest virtual world platform is SecondLife. It was developed in Linden Labs and launched in 2003. In the first quarter of 2011 it had an average of 794,000 repeat logins by the platform's users (SecondLife, 2011). It generated US$100 million in

the year which made it a significant independent economy, trading in Linden dollars which have a direct exchange rate with US dollars. By 2012, SecondLife became involved in game playing through a product called Linden Realms and quite clearly they see a huge importance for gaming in the future development of SecondLife. In August 2012 they announced their collaboration with Steam, the gaming platform which attracts 40 million players with up to 4 million playing at any one time. There's an obvious potential for the development of language learning games within this type of environment which will be discussed more fully in the conclusions section of this thesis (see Chapter 6).

Sykes, Oskoz and Thorne (2008) used a specific type of virtual world, a synthetic environment, in order to teach specific pragmatic features associated with Spanish requests and apologies. Synthetic environments are artificial representations of reality, the flight simulator being the first and best known example. Students were introduced to synthetic immersive environments (SIES), which had explicit learning aims and relevant interactional contexts. The developers claim to have created a collaborative space in which learners are at the centre of their own learning. Amongst the reasons Sykes (2008) found for the usefulness of the platform were the possibility of multi-modality processing for learners and multi-level feedback with minimal instructor interference. This type of project offers new directions for the development of CALL materials.

Three dimensional multi-user virtual environments (3-D MUVE), such as those created by Barab, Arici and Jackson (2005), offered great but largely unrealised possibilities in recent years. While the advantages for 3-D MUVE are obvious it is claimed that they have become too complicated for many language teachers to use and also too costly. A 16 acre island in SecondLife (owned by Linden Lab) typically costs in the region US$1,675 with maintenance of US$295 per month (Cooke-Plagwitz, 2008). Many institutions may have neither the technical nor financial resources to participate.

2.5.4 SLA and chatbots.

Another less widely used concept pertaining to the virtual world, in the context of second language acquisition, is that of chatbots. Chatbots are described as "computer programs that interact with users using natural languages" (Abu Shawar & Atwell, 2007a, p. 29). Chatbots come in a variety of forms, many having avatars with greater or lesser similarities to a human being. They are normally internet-based and are designed to simulate an intelligent human conversation. Some are entirely text-based while others have an integrated TTS facility (see Section 2.6 for further detail on TTS) and possibly a voice recognition component.

Initially chatbots were developed mainly for fun and used a technique which took keywords from the input and matched them to predetermined output patterns. A typical example of this was ELIZA (Weizenbaum, 1966) which was developed primarily for fun rather than having any real practical use. Since ELIZA, however, more sophisticated forms of Artificial Intelligence (AI) architectures have been developed and their use in the domain of business and commerce is mushrooming. Van Lun (2013) of chatbots.org reviews a list of 11 speech synthesis chatbots launched between May and October 2012. These are all developed by commercial organisations who use them primarily to facilitate their sales environment. (See Rudo (2011) for enthusiastic, even if somewhat exaggerated, claims for the usefulness of chatbots). More recent AI architectural structures, such as those used in Jabberwacky are such that they learn from the interactions they have with visitors (Fryer & Carpenter, 2006). Jabberwacky has stored several million conversational patterns and can learn any language given the appropriate input. It has the ability to correct common spelling and grammar errors.

Chatbots are still a very new addition to CALL or to education in general. There is not a great deal of literature on their use and there are few serious evaluations of their usefulness as a component of CALL (Parker, 2005). Their use for a number of educational purposes such as teaching mathematics (Knill, Carlsson, Chi, & Lezama, 2004) or training medical students (Kerfoot et al., 2006) has been advanced, but while

initial trials seem promising there has not been a widespread acceptance or use of chatbots in any of these areas.

Chatbots are getting some attention in CALL but their use to date has been more at developmental and at trial stage and their value remains a matter of speculation. I am unable to find any large scale CALL program where chatbots have an inherent role. As Fryer and Carpenter (2006) note, there would appear to be no chatbot designed from the ground up specifically for CALL.

It is postulated that chatbots may have a significant contribution to make to CALL in that they allow learners to participate in a life-like social interaction where meaning can, to a greater or lesser extent, be negotiated, thus encouraging language acquisition (Cooke-Plagwitz, 2008). At the present stage of development, however, meaningful interactions are limited in range and confined to restricted topics in order to enable students to practise specific linguistic structures and forms (Coniam, 2008). They tend to be used as a means of giving learners a low cost way of practising particular language skills. Lehtinen (2008, p. 3) sees them as "a scaffolding device for those who are still at the early stages of their foreign language development or are terrified of interacting with people". The learners' privacy in interacting with the computer may reduce stress and thereby facilitate language acquisition (Wang, 2008). Fryer and Carpenter (2006) offer six advantages for the use of chatbots. Among these advantages are: the relative anonymity of the learner in interacting with the computer; the computer's ability for endless repetition where a human interlocutor would get impatient; the newness of chatbots for learners make them more interesting; and learners may experiment with language structures and get a reaction from the computer. On the other hand, most chatbots are unable to give an appropriate corrective reaction to incorrect grammar or spelling errors. They lack naturalness in their responses and sometimes the responses may be inappropriate (Jia, 2004). Some chatbots have incorporated a speech recognition system but these systems require a lot of improvement before being satisfactory in a high level human-computer interaction (Robin, 2007).

It is claimed that chatbots offer the opportunity for computer-simulated natural language social interaction which can open new horizons for spoken language teaching/learning (Sha, 2010). To date, however, their use has been confined to being a tool to "practise the language and enjoy chatting" (Abu Shawar & Atwell, 2007a, p. 38). In the six years that have passed since the Abu Shawar and Atwell review, to the time of writing the present thesis (2013), not a great deal of progress has been made in the incorporation of chatbots into CALL. Georgila, Black, Sagae, & Traum (2012) are still at the point of speculating on design features of chatbots including the merits of synthetic voices. If chatbots are to become a serious part of CALL, then considerably more research is needed in their development.

The review of the literature outlined above suggests that chatbots may be an effective tool to combine communicative language learning, the concept of engagement with a task, and the use of TTS synthesis. For this reason a chatbot as a CALL activity (*Taidhgín*) is being developed as one of the three platforms for the present study. A detailed description of its development is given in Chapter 3.2.3.

2.5.5 Characteristics of CALL games.

In the following section the characteristics of good language learning games are discussed. The possibility of language development in a virtual world context will be discussed in the concluding section of this thesis (see Chapter 6.9).

Language learning games have certain features which set them apart from general games in that they aim to enable participation in the game by focusing the player's attention on the linguistic forms necessary for the achievement of the defined goals. They must therefore achieve a balance between focus on language form and language function. One can only accomplish a specific function by acquiring a specific language form. There has been relatively little research on this balance. DeHann, Reed and Kuwada (2010) examined the extent to which video game interactivity would help or hinder the noticing

and recall of second language vocabulary. Eighty randomly selected Japanese university graduates were paired and randomly assigned to either playing an English-language music video game or watching it being played while they remained inactive at a separate monitor. The watchers recalled significantly more vocabulary from the game than did the players. This is put down to the extraneous cognitive load induced by the interactivity of the game (there is further discussion on cognitive load theory in Section 2.5.6.1 below). The attention of the players became diffused by the demands of the interactivity suggesting that the balance between form and function was not optimal.

The research on cognitive overload and language games, though relatively sparse, is nevertheless informative. There has been some research on the value of annotation of film presented to learners in the target language by way of adding subtitles, captions or audio prompts. The results of research are varied. It was believed that multiple modalities of input would improve language acquisition. Paivio's (1986, 1991, 2007) theory of Dual Coding claimed that verbal and non-verbal stimuli are processed by two different systems but that the two systems interact and the activation of both systems simultaneously leads to better recall. Yanguas (2009) found that all multimedia gloss groups noticed and recognised significantly more target words than a control group which did not have access to glosses. She also found that those who had a combination of glosses outperformed those with single glosses. Similarly, Winke, Gass and Sydorenko (2010) found the captioning of videos to be more effective than non-captioning in respect of vocabulary acquisition and comprehension. Interviews with the learners suggested that they used the captions to direct their attention and improve processing by helping them to analyse the language input. Similar results were obtained by Al-Seghayer (2001) who found that video and text led to better vocabulary learning than annotations that combined pictures and texts. Danan (2004) suggested that subtitles be confined to very difficult material, otherwise the use of captions in L2 should be used.

Sydorenko (2010) suggests that different types of video input seem to provide different benefits. She found that video plus audio plus captions left learners with a greater written

ability than their aural recognition ability while the group exposed to just video and audio stimuli had a higher aural recognition ability.

The multiple modalities concept is brought into question by Sweller's (2005) redundancy principle of cognitive overload. He claimed that redundant material slows down information processing and learning. Attention capacity is limited and learners need to use it selectively. If learners' attention is divided by three different modalities they may experience cognitive overload (Robinson, 2003; Wickens, 2007).

It is difficult to make any generalisations from apparently contradictory research results. Frequently the groups tested had very small numbers and taking generalisations from them is unsafe. We may take their findings as being indicative rather than conclusive and will suggest broad general outlines for the design of CALL games.

2.5.6 Features of CALL game design.

Learner engagement is essential to learner success (Lim, 2008), where engagement entails mindfulness, intrinsic motivation, cognitive effort and attention (Herrington, Oliver, & Reeves, 2003). This suggests that game playing is a holistic experience and that one's view of particular aspects of the game, for example graphics or plot, is influenced by one's disposition towards the whole of the game, i.e. the extent to which the game induces 'flow'. The concept of wholeness and its importance in the context of technology-enhanced learning environments in schools is referred to again by Lim and Chai (2008). For the purposes of the present study, this may have implications for learners' judgments of the quality of the synthetic voices. It would seem likely that since the voices are being judged in the context of a game-like platform that one's overall experience, sense of 'flow', or general disposition towards the game, will influence one's attitude towards each element of the platform. One would therefore expect a high correlation between judgments on elements of the platform such as graphics and plot and judgments of the intelligibility, quality and attractiveness of the synthetic voices. If one has a high opinion of the graphics and plot, for example, it is likely that this will lead to a

greater attraction to the game and a subsequent increase in rating for the intelligibility, quality and attractiveness of the synthetic voices. The opposite may also apply.

The most significant strength of digital games is their potential to motivate the player or learner in the case of serious games. In her review of literature, Dickey (2011) concluded that intrinsic motivation, curiosity and plausability were first supported by the game-like environment and then sustained by the narrative. This entails designing games with concrete focused activities having:

(a) the ability to engage the learners' interest

(b) a primary focus on meaning

(c) a game which is possible to complete rather than being open-ended

(d) a judgment of the outcome in terms of success or failure

(e) a clear relationship with real-world activities, i.e. tasks are seen as having real-life significance.

2.5.6.1 Cognitive load theory.

In considering the design features of serious games one needs to give considerable attention to cognitive load theory (CLT). The theory has been developed substantially since the 1980s and tries to base instructional design on postulated interactions between information structures, on the one hand, and human cognition on the other (van Merrienboer & Sweller, 2005). It relates to the mental effort required for successful completion of a task (Pillay, 1994). CLT assumes a very limited working memory which can deal with new information for no more than a few seconds unless it is refreshed by rehearsal. This operates differently from long-term memory which holds cognitive schemata and can alter the characteristics of working memory and greatly expand its capacity by giving it reference points. It is further postulated that visual and auditory working memory are partially independent of each other. If, for example, multiple

sources of information are all presented in a visual format (video, text, diagram) they are more likely to overload the visual processor than if the written material is presented in spoken form, allowing for some load shift from the visual to the auditory processor. Kataoka (2008) reports a number of Japanese studies which showed that words with TTS voices helped junior high school students to learn and memorise more English words than did their counterparts who were without TTS help (Sasaki, 2006). Kataoka (2007) found that words and phrases with TTS sounds helped university entrance students learn and memorise them more easily. Yoshida (2008) found that TTS voices with animation enhanced the ability of Japanese university students to learn English. Cognitive overloading on either the visual or the auditory processor causes significant delays in problem solving as well as failure to apply correct rules and ultimately a breakdown in participation in the task (VanLehn, 1999). Zheng, McAlack, Wilmes, Kohler-Evans, & Williamson (2009) add the concept of manipulative learning to perceptual (visual and auditory) learning. Manipulative learning refers to kinaesthetic stimulation or object manipulation associated with achieving a certain goal. Individuals who are kinaesthetic learn best with an active "hands-on" approach (Pourhossein Gilakjani, 2012). This adds a third learning channel to the auditory and visual channels onto which the cognitive load may be shifted. Zheng et al. (2009) found that providing learners with a manipulative function in multimedia learning reduces their cognitive load and improves their problem solving ability.

2.5.6.2 CLT and computer games.

Mayer and Moreno (2003) adapted a number of the key principles of CLT which they regarded as essential to the design of serious games. They identified three facets of the cognitive processing involved in learning through gaming, viz., (a) essential processing (making sense of the material being presented) (b) incidental processing (the non-essential aspects of the material) (c) representational processing (holding verbal or visual information in working memory). They see the following steps as necessary for avoiding cognitive overloading:

1. Avoid channel overloading with essential processing demands. The material being presented should not be over rich and demanding on the learner. Material should be divided between sensory channels by providing verbal and visual information simultaneously.
2. Segment essential information or give prior preparatory information. Avoid presenting rich material at a fast pace so that learners may organise words and images from one segment before being confronted by the next segment.
3. Avoid extraneous materials which cause overload by incidental processing. They found that learners did better with concise narrated animation than they did with embellished narrated animation, e.g. background music or video clips showing irrelevant material.
4. Avoid presenting material in a confusing way, e.g. too much material on the screen, text separated from graphics such that learners are confused by having to divert their attention from one area of the screen to another.
5. Present visual and auditory material simultaneously rather than successively so as to avoid the learner having to hold auditory information in working memory while dealing with the visual.

Other researchers have pointed to further design issues which help to reduce cognitive load. Keller, Gerjets, Scheiter and Garsoffsky (2006) found that three-dimensional designs were associated with higher cognitive load than were two-dimensional ones due to their additional complexity. Van Schaik and Ling (2003) found that the more familiar the link colour and screen page design for web pages, the less was the cognitive load and they suggested that designers for instructional material should stick to standard web form, screen layout and navigation in order to maintain students' browsing familiarity. Mikk (2008) found a strong relationship between length of sentences and working memory overload and this affected reading comprehension.

2.5.7 Criteria for the design of CALL games.

The design of serious games then must steer a middle course between a number of apparently contradictory demands.

1. CALL games must be predominantly task-based with focus on meaning and on the use of the target language in order to achieve clearly defined outcomes (Ellis, 2005).
2. Learners need to develop metalinguistic awareness by playing the game. This factor can be built into the game by varying the subtlety of the language choices given to the player. This may vary from totally inappropriate utterances for a particular context to slight pragmatic variations in other utterances.
3. All elements of the game should have a playful spirit (Purushotma, Thorne, & Wheatley, 2009).
4. The game screen needs to have enough information to clarify the situation being depicted but should not contain extraneous material which may be distracting or leading to cognitive overload. Additional metalinguistic notes such as grammar notes, dictionary items, etc. may be included by way of hypertext where the learner can deliberately seek out help if required. CALL games are not in competition with the commercial gaming market as the criteria for the success of each are quite different (Kickmeier-Rust, Albert, Hockemeyer, & Augustin, 2007).
5. CALL games need to have clearly defined criteria for success for the player/learner such as finding the answer to the puzzle, reaching the required destination, acquiring the necessary information, etc. Recognition for partial success may be given by way of a points system which credits correct or acceptable moves.
6. CALL games must maximise 'flow' or the engagement of the learner with the task being presented, yet they must not be so engaging on the activities side that the language learning aspect of the game is overlooked. The fun involved in

deliberate failure or random selection of responses should not be greater than that got from the correct response.

7. CALL games should have multimodalities with visual, auditory and manipulative input to the game. Visual and auditory material should be presented simultaneously in order to help recall.
8. The narrative of the game needs to be culturally legitimate and to have a credibility for the player/learner.
9. CALL games should avoid cluttered screens or screens with an unfriendly layout which divide the player/learner's attention.

2.5.8 Summary.

This section has examined the concept of computer games as an aid to learning, and as an area of CALL which has remained relatively underdeveloped to the present time. It is noted, from a connectionist viewpoint, that the functioning of computer games can be designed to resemble the functioning of the human mind. If the architecture of a computer game can be such that it resembles that of the functioning of the human brain then optimal learning conditions can be created. Attractive as this theory may be, it remains unproven. A more realistic field of endeavour is examining what features go to constitute 'good' games. The essential element for a game is its ability to hold the attention of the player. The term 'flow' is used to refer to this ability.

Games which have some educational intent are referred to as serious games and a distinction is made in the literature between these and games which have purely an entertainment value. The essential features of both are quite similar and, it is argued, making such distinctions is unhelpful for the future development of computer gaming.

One effective context in which gaming can be exploited is that of virtual world platforms, such as SecondLife. The idea of using CALL games in such a virtual world has not yet been exploited to any great extent.

A related idea for CALL platforms is the use of chatbots or virtual dialogue partners which are supported by artificial intelligence (AI) systems. Chatbots offer considerable potential for CALL platforms but, as yet, this potential remains largely unexploited.

In considering the characteristics of good CALL games, one must try to match the game features to the human cognitive processing system. The features of the game must be such that they allow for maximum cognitive functioning. If the cognitive processing becomes overloaded then the attention of the player becomes diffused and learning will be inhibited.

Based on the available literature, nine key features of good CALL games have been identified. Combined, these features aim at creating games where the player has high engagement with an activity; where the activity itself is purposeful and has cultural validity; and where the demands of the game match the cognitive processing ability of the player. These key features become the foundations on which the platforms being developed in the present project are based.

2.6 Text-to-speech Synthesis

2.6.1 Overview.

In considering CALL programs it seems self-evident that the spoken language should be regarded as an essential dimension of such programs. Most modern language teaching programs place emphasis on the learner's oral/aural communicative skills. Traditionally, the spoken texts used were pre-recorded.

In the context of CALL programs, which are highly interactive and which are driven by AI systems, the interactions between the player and the computer become relatively unpredictable. Novel utterances may be produced by the player which require novel responses from the computer. In this case, the pre-recording of material is no longer possible. In an ideal world, one would like to have both speech recognition and TTS capabilities built into the CALL system. The development of speech recognition facilities that can deal with the speech of second language learners with various degrees of proficiency is still a long way off. In the case of Irish, speech recognition facilities are not available, and so this aspect falls outside the parameters of the present study.

The use of synthesised speech in CALL is still at an early stage. The present section outlines the general issues relating to synthetic speech and CALL. It then goes on to provide background on the development of the first TTS voices for the Irish language. A description of the systems used to produce these voices is important, as part of the overall aim of the present project is to test their acceptability in newly developed CALL platforms.

2.6.2 Text-to-speech in CALL.

Text-to-speech (TTS) synthesis has been exploited primarily to address the needs of enterprise and telecommunications systems, such as automated answering services. More recently, its use has been extended to desktop speech systems, computer voice interfaces, audiobooks, dictionaries, satellite navigation, etc. TTS did not always arouse a great deal

of interest among second/foreign language teachers because of poor prosody, lack of clarity and consistency (Sha, 2010), particularly in the case of early systems. A review of the literature on the use of TTS synthesis for educational purposes showed the 1990s as the decade which promised most hope for the future of TTS systems in language teaching/learning. Stratil, Burkhardt, Jarratt and Yandle (1987) were positive for the future of TTS synthesis in the teaching of Spanish grammatical constructs as far back as 1987. They noted a positive reaction from students and teachers. They saw the need for providing easily accessible tools for teachers so that they could exploit TTS potential. Klatt (1987), who was a major contributor to the development of TTS synthesis, saw it as having the possibility of being helpful for people with dyslexia. Atkinson (1974) suggested that a TTS system may have value for people with dyslexia in that it would save their embarrassment of getting help from a real teacher. This claim may seem a bit spurious for modern western culture where dyslexia is no longer stigmatised. There has been an increased exploitation of TTS as an aid for learners with dyslexia (Elkind, Cohen, & Murray, 1993; Greene, 2006). Gelan (2010) suggests confining the use of TTS to users with reduced sight or with auditory learning preferences and sees few CALL applications using TTS other than for talking dictionaries. Olson and Wise (1992) suggested that it would have a role in improving students' ability to sight-read and decode. Hecker, Burns, Elkind, Elkind, and Katz (2002) found that when texts were presented through visual and audio channels simultaneously they allowed students with attention disorders to concentrate for longer periods of time. A number of researchers have examined the role of TTS systems in developing reading skills. The overall result of these studies suggest TTS systems are useful for students with reading difficulties, of very little use to the average student, and may inhibit more advanced readers (Disseldorp & Chambers, 2002; Drezek, 2007; Elkind, 1998; Leong, 1995; Montali & Lewandowski, 1996).

As mentioned, TTS synthesis has not been widely used to date in CALL (Gupta & Schulze, 2012). One can only surmise that this may be due, in part at least, to a lack of

satisfaction amongst language teachers with the quality of the synthetic voices that are available. Keller and Zellner-Keller (2000) report synthetic voices being insufficiently expressive and lacking in any form of human emotion. The range of synthetic voices is very limited within any language, usually confined to a small number of male and female voices with few, if any, children's voices. Delogu, Conte and Sementina (1998) found that listening to and comprehending synthetic voices was more difficult than comprehending natural voices. Higgins (1988) was quite clearly referring to a less well developed TTS system when he concluded that one should not bother using synthetic speech as a model for learners because it cannot account for the complexity of human language. Many of these criticisms pertain to older TTS systems, and it should be noted that there have been huge advances in the TTS technology in recent years though some claim that the basic criticism still holds (Handley, 2005). However, it goes without saying that the quality of synthetic speech is highly variable, and that, at the top end, it is very good indeed. Nonetheless, in the case of languages other than English and major world languages, such quality may not be available.

The applications that are in use, exploiting TTS, are mostly confined to talking dictionaries, talking texts, dictation exercises, and practice of individual and combined phonemes (Handley & Hamel, 2005). Recent use of TTS systems has also included the development of individual language skills, such as listening comprehension or pronunciation skills, amongst learners of English, primarily. What little evidence there is, is encouraging for the continuing development of TTS as a facet of CALL. It was found, for example, that a TTS system helped Turkish learners of elementary English improve their pronunciation (Kılıçkaya, 2011). González (2007) concluded that her class of architectural students who were studying English as a second/foreign language benefited from TTS applications by improving their pronunciation skills and by being encouraged to become more autonomous learners. She found that when her students listened to the text of the professional papers which they needed to study, while reading them simultaneously, they had better retention of new vocabulary. They also listened to texts,

produced by themselves in written form and this improved the presentations of their papers to colleagues. Synthetic speech seems particularly useful for this type of exercise as the texts are unpredictable and it would not be feasible to have a native speaker record such a wide variety of texts.

I have not been able to find any examples of a TTS system being used as an integral part of a serious CALL game with the exception of 'The Language Trap', developed by the Knowledge and Data Engineering Group in Trinity College, Dublin (Peirce & Wade, 2010). The main reasons for this are outlined above and relate to language teachers' and program developers' dissatisfaction with the quality of the TTS systems that were available to them. It is still believed that for TTS synthesis to be of use in CALL programs, the voices must be very close to that produced by a human. A high degree of naturalness is necessary if the aim of the lesson or module is expressed in terms of metalinguistic outcomes, such as accuracy of pronunciation or vocabulary retention. If language teaching aims have changed such that the focus of activities is on meaning rather than linguistic form, then the criteria for judging the adequacy of TTS systems may also be changed. We can now ask if the TTS tool is adequate for purpose or 'authentic enough' for the purpose it is intended to serve (Kılıçkaya, 2006). Handley (2009) asks 'Is text-to-speech synthesis ready for use in computer-assisted language learning?' and she concludes that the best TTS synthesis systems are ready for use in applications where they 'add value' to CALL. She was referring to the use of TTS synthesis as (1) a reading machine, (2) a pronunciation model and (3) a conversational partner.

CALL has not, to date, exploited the potential of TTS in language learning. Perhaps the expectations of those involved in course design or in designing CALL programs have exceeded what TTS systems can deliver in that it is unlikely that a computer can ever truly express the range of human emotions with the full human range of prosodic and tonal features. To wait for computers to master these features is to despair of the use of

TTS synthesis in CALL games at any time in the foreseeable future. The 'fitness-for-purpose' proposal being forwarded here accepts a standard of TTS synthesis which is less than perfect but which will encourage greater interactivity between the learner and the language game and offer an engagement in language learning which is outside the range of currently available teaching materials.

As mentioned in Section 2.2.8, using CALL in the teaching/learning of endangered languages is uncommon. TTS synthetic speech for minority or endangered languages is also unusual and, indeed Irish (and the *ABAIR* project) is in quite a unique position in this regard (see Auclerc et al. (2000) for synthetic pronunciation of Breton words in dictionary format). The integration of synthetic speech into interactive CALL platforms is rare as a language teaching technique, even for the major world languages. I am unaware of any such approach to the teaching of minority or endangered languages.

The use of synthetic speech for Irish offers learners the possibility of having access to native-like speakers. Since Irish is treated as an endangered language and, given its current relatively weak position in Gaeltacht areas (see Section 2.2.1), it is likely that the number of native speakers will continue to decline. In such a scenario access to native Irish language teachers will become more rare and synthetic voices may have increased importance.

2.6.3 Developing a TTS synthesis system for Irish: the *ABAIR* project.

The development of an Irish language TTS synthesis system and speech technology resources has been ongoing for a number of years in the Phonetics and Speech Laboratory of the School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin, under the direction of Ailbhe Ní Chasaide. This research initiative stemmed from groundwork carried out as part of the WISPR (Welsh and Irish Speech Processing Resources) project, an EU collaboration under the INTERREG IIIA

programme, which involved collaboration with the University of Wales, Bangor, with input also from University College Dublin, Dublin City University and Institiúid Teangeolaíochta Éireann. The aim of WISPR was, inter alia, to produce corpora along with speech and linguistic resources which would underpin an Irish language synthesis system. The project continued with the research initiatives *Cabóigín/Cabógaí*, funded by Foras na Gaeilge; *CabairE*, funded by An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta (COGG); and continues with the current *ABAIR* project, funded by An Roinn Ealaíon, Oidhreachta agus Gaeltachta (The Department of Arts, Heritage and the Gaeltacht). Over the years the remit of the research has broadened to encompass applications that target disability, access and educational exploitations for learners of Irish. It is in the latter context that the present work has been developed. For more information on the *ABAIR* development, see the information on the *www.abair.ie* website and also (Ní Chasaide et al., 2006, 2011; Ní Chiaráin & Ní Chasaide, 2012; Ó Raghallaigh, 2009; Prys et al., 2004; Williams, Prys, & Ní Chasaide, 2005; Wogan et al., 2006).

The synthetic voices developed in this research are available for use on the *www.abair.ie* website: voices for two dialects, Gaoth Dobhair (Donegal) and Ráth Chairn (Conamara), along with some of the underpinning linguistic resources (described in Section 2.6.4), have been made available on the website, and a Munster voice is under development.

The multi-dialectal approach to the development of a TTS system for Irish was adopted as there is not a single spoken standard dialect of Irish. The dialect specific voices confer the benefit of authenticity and, in practical terms, any educational tools will need to incorporate the dialect diversity of modern Irish. Furthermore, in the context of the decreasing numbers of native speakers of various dialects (Ó Ghadhra, 2001; Ó Giollagáin, Mac Donnacha, Ní Chualáin, Ní Shéaghdha, & O'Brien, 2007), the production of authentic dialect specific TTS is important for the future.

Within the *ABAIR* project, synthetic speech is being developed using two different methodologies, namely unit-selection synthesis (Black & Campbell, 1995; Hunt & Black, 1996) and the hidden Markov model (HMM)-based speech synthesis system, HTS (Black, Zen, & Tokuda, 2007; Zen et al., 2007). The quality of synthetic speech produced with these two techniques can be rather different, and each method has particular advantages and disadvantages. A brief description of each system is given below in Sections 2.6.3.1 and 2.6.3.2.

The distinction between the various methods of producing synthesis is worth noting in the present thesis since one of the three platforms (*Taidhgín*), described in Chapter 3.2.3, uses HTS synthesis while the other two platforms (*Digichaint* and *Fáilte go TCD*) use unit-selection synthesis. The differences, as perceived by the participants in the study, between the output of the two systems will be noted in the results section.

2.6.3.1 Unit-selection synthesis.

Unit-selection synthesis is based on recordings (a large read corpus) of real human speech. It is a form of concatenative speech synthesis where potentially minute segments of recorded speech are pieced together in order to produce synthetic speech output. This is one of the most popular techniques in use today, particularly by commercial companies, as it generates the most natural sounding speech output. One of the main benefits of this synthesis system is that it preserves the natural voice quality of the speaker. One major drawback of this method of synthesis is that it lacks flexibility for modifying the characteristics of the voice. Another disadvantage of unit-selection synthesis is that there can be concatenation errors and sudden discontinuities and, when extreme, they can seriously reduce the quality of the output. These poor concatenations can be caused by different factors. For example, there may be missing contexts (e.g. sounds in particular phonetic contexts) in the speech recordings on which the synthesis is based, with the result that the system may sometimes fail to find suitable sounds to concatenate. Errors in the processing of the speech corpus may also give rise to problems, such as mistaken segmentation, labelling of the sounds, etc.

The general framework on which the *ABAIR* system is built is Festival 1.95 which is a version of the Festival Speech Synthesis System (Clark, Richmond, & King, 2004) developed at the Centre for Speech Technology Research (CSTR) at the University of Edinburgh. Festival offers a general framework for building speech synthesis systems. It is an open source system from which a number of speech engines are available for the development of new voices using unit-selection synthesis.

Although robust, natural sounding voices have been created using this system, one other disadvantage of the current open source implementation of unit-selection synthesis in *ABAIR* is the large size of the corpus/database used in the synthesiser. A large memory is needed to store it - the process of generating synthesis involves searching the corpus to locate the most suitable speech segments, and choosing the best candidates. This process necessarily takes time, and so the generation of the speech output is slower than it is for HTS synthesis where only parameter values are stored. So, where rapid real time responses are needed, HTS synthesis is the optimal choice.

Unit-selection synthesis was used as the characters' voices in the platforms *Digichaint* and *Fáilte go TCD*. As the utterances the characters would pronounce were pre-prepared in these platforms, it was envisaged that any concatenation errors that might arise could be mitigated somewhat by slightly modifying the text input where necessary. It was decided that the lack of buzziness (associated with HTS synthesis, see below) and more natural voice quality better suited these applications.

2.6.3.2 HTS synthesis.

HTS is very different from the unit-selection method. In HTS, acoustic models of speech are trained using machine learning technology similar to that used in speech recognition. This method uses a speech corpus also, and involves machine learning of the acoustic characteristics of individual phonemes of the corpus in the various contexts in which they

occur (the context thus includes the phonetic neighbourhood, the phoneme's position in the phrase, whether it is stressed, etc.). At synthesis time, the various parametric descriptions of speech (to do with pitch, duration and spectral characteristics) are predicted from the input text, and these sources of information are used to construct the synthetic speech waveform. This method of synthesis generates speech parameters on the basis of the models constructed from the training data. Synthetic speech is then created using these generated parameter contours.

HTS and unit-selection synthesis both rely upon a corpus of recorded speech. The advantage of HTS over unit-selection synthesis, however, is that to generate speech it has far smaller requirements in terms of computational resources. This has practical implications such as being able, for example, to store the system on mobile devices which would have low-storage capacity. The development of a HTS synthesis system for Connacht Irish was necessary because the unit-selection synthesiser did not fully meet the demands of specific groups who wished to use the synthetic voices, such as visually impaired users/learners of Irish (rapid speech generation is essential for applications such as screenreaders for the visually impaired). Concatenative approaches can generate speech at a faster rate but only to a limited degree. For the dramatic increase in the speed of the speech output often required by blind screenreader users the fully parametric approach of HTS is more suitable.

A drawback of the HTS system is that although it produces good quality synthesis, the quality of the voice is not as natural as in high quality unit-selection synthesis – it has a 'buzzy' and somewhat 'robotic' quality to the speech output due to the strong harmonic structure in the excitation signal (Zhang, Tao, Jia, & Wang, 2008). Separate to the problem of 'buzziness' there can also be an issue with over-averaging of the acoustic models and this can lead to a 'muffled' quality in the speech output (Toda & Tokuda, 2007). It must be noted, however, that the quality is improving rapidly and, furthermore, the quality of the output is stable as there are no sudden discontinuities due to poor concatenations as can happen with unit-selection synthesis. It is therefore preferable to

use HTS synthesis in certain applications such as screen readers as its output tends to be more predictable insofar as HTS will not produce sudden discontinuities, bad concatenation errors or incomprehensible output.

The focus of much of the academic research in speech synthesis is currently being directed at HTS synthesis and improvements are being made to the system on a continuous basis. Current academic developments are concentrating to a large extent on voice modification using HTS synthesis (Veaux, Yamagishi, & King, 2012; Yamagishi, Kobayashi, Nakano, Ogata, & Isogai, 2009). This would have immediate significance for applications such as those developed in the present project where virtual worlds could be populated with characters (all sounding completely different from one other) – all derived from a limited number of TTS synthetic voices.

HTS was used as the method of generating TTS synthesis in the *Taidhgín* platform because of the necessity to produce responses almost instantly, giving the feeling of a real time conversation. For a fuller discussion on this platform see Chapter 3.2.3. By evaluating both unit-selection synthesis and HTS synthesis across the three platforms, one gets a deeper insight into user preferences and the relative strengths and weaknesses of each method in various different contexts.

2.6.4 Building *ABAIR* voices.

An important task in building any synthetic voice is to select a suitable speaker. He/she needs to be a native speaker of the dialect who is a fluent reader, has a suitable quality to his/her voice (excessive creakiness or breathiness may be avoided, for example), and is comfortable with the task of reading from a screen in a recording studio. A female Ulster-dialect speaker, a male Connaught-dialect speaker and a female Munster dialect-speaker were selected for the *ABAIR* project.

Texts for the recording, which were selected from literary works originally written in the dialects of the speakers, were chosen. Some editorial work was done in order to make these texts more suitable. Recordings were carried out over several months, adding up to a total of about 50 hours. Care was taken to ensure consistency in recording quality by attending to details such as continuity of practice in the distance between the microphone and the speaker; consistence in the times of day during which the recordings were made and the exclusion of extraneous noise or interferences. Two people were present during the recordings to take care of the running of the software and the feeding of the prompts to the speaker as well as taking notes on pronunciations with a view to building up an 'exceptions lexicon'.

In order to generate synthesis the text has to be prepared so that it can be fully represented in orthographic form. This involves the generation of orthographic forms of numerals, abbreviations and other symbols. Subsequently, the orthographic forms need to be converted into strings of sounds. The sound system of Irish is large and complex (see Ní Chasaide (1999) and Ó Raghallaigh (2012)), and this has implications for the letter-to-sound rules and for the size of corpus needed (as every sound combination of the language has to be represented). Irish letter-to-sound correspondences, though complex, are reasonably regular but any pronunciation the speakers had of individual words that did not fit within the general letter-to-sound rules that describe the dialect, were noted and stored explicitly in what was called an 'exceptions lexicon'. The list includes common phrases that exhibit features of fluent running speech, such as occurrences of reductions, deletions, and cases where the phrase is not uttered as a sequence of individual words but as one phonetic word ('cén chaoi a bhfuil tú?' -> c e: x i: vʸ i lʲ t̪ʸ u:). In an effort to ensure that the synthesiser would sound as natural as possible, all these features were programmed explicitly into the system.

The corpus was then annotated, a process which involved segmenting it into sentences, words, syllables and phonemes. The process was automated using a technique known as 'forced alignment'. For more details on this technique see Yoshimura, Tokuda, Masuko,

Kobayashi and Kitamura (1999). It involved feeding the phoneme strings of the text along with its corresponding sound files into the aligner and using a software toolkit to carry out the alignment. The Hidden Markov Model Toolkit (HTK), a toolkit for building and manipulating hidden Markov models (Young et al., 2009), was used for the present implementation of the *ABAIR* project. Although in practical terms it would be near impossible to annotate the full corpus without this automated technique, it is still far from perfect and much time is still needed in order to manually correct the annotation so as to ensure sounds are labelled correctly. The better the annotation on the corpus, the better the synthetic speech output.

Other information was also stored about each segment of speech, such as fundamental frequency, duration and context. The context refers to the segments which surround the target segment – this is especially important to capture coarticulation, as the acoustic characteristics of every sound are greatly influenced by the context in which they occur. Context also encompasses the prosodically relevant contextual information, such as whether the sound in question belongs to a stressed syllable or not and its position (word-initial, word-final, phrase-initial or phrase-final).

2.6.5 Currently available resources.

To date, the approximately 50 hours of recorded speech has been analysed and annotated for the *ABAIR* TTS synthesis system. This includes c. 30 hours of Donegal Irish (female speaker), 5 hours of Conamara Irish (male voice, actually from Ráth Chairn) and 15 hours of Kerry Irish (female voice). Along with the synthetic voices, the following resources are available to the public on the *www.abair.ie* website:

2.6.5.1 Phonetisation tool.

The phonetisation tool allows one to view a phonetic representation of any word/utterance entered into the system. This phonetic transcription refers to a precise description of the sounds which are to be found in a given language. A description which distinguishes phonetisation from ordinary orthography is available for the general user on

the *www.abair.ie* website. Note that the letter-to-sound rules are complex in Irish: a specific letter/sequence of letters can map to many sounds: for example orthographic ***uí***, ***aoi***, ***ao*** and ***í*** can all represent a single phoneme /i:/ (as in the words ***luí***, ***maoigh***, ***gaoth***, and ***rí*** in Donegal Irish). Conversely, a single orthographic form can represent different sounds: for example, the sequence '***adh***' of ***adh***mad and rug***adh*** are pronounced with differing sounds, /aɪ/ and /u:/ respectively (in the Connaught dialect). The phonetisation tool is useful to language learners who are familiar with the IPA phonetic alphabet, in that the orthography-to-sound mapping is made explicit.

2.6.5.2 Text normalisation tool.

This transforms non-orthographic forms (such as €, abbreviations, numerals, etc. into orthographic forms. To take the example of the numerical system, note that it is relatively complex in Irish. '2' can be expressed as *dó*, *dhá*, *beirt*, *dara* / *dóú* depending on the context in which it arises. Rules are written explicitly into the synthesiser in order to handle this and context needs to be taken into account. Other examples of rules that feature in this module include the symbols *1/2* as ***leath*** (half), *€2* as ***dhá euro*** (two Euro), *2 chailíní* as ***beirt*** *chailíní* (two girls), *1 nó 2* as *ceann nó* ***dhó*** (one or two) and *seomra 2* as *seomra a* ***dó*** (room two). Text normalisation is useful as a teaching tool for certain lessons based on the numerical system, such as learning to tell the time, explaining the difference between different types of numbers, etc.

2.6.5.3 Part-of-speech tagger.

One can also access a part-of-speech (POS) tagger on the *www.abair.ie* website. POS tagging involves a grammatical analysis of each word that is in a given sentence. A label is given to the word denoting the word class (e.g. noun, verb, adverb, pronoun, etc.) as well as the grammatical form (e.g. person, number, gender, tense, mood, case).

Input: **Tá** –> Label: *verb, present tense*; (Translation: am)

Input: **mé** –> Label: *pronoun, 1 person, singular*; (Translation: I)

Input: **sásta** –> Label: *adjective;* (Translation: happy/satisfied)

Figure 2.4. An example of Part-of-Speech tagging applied to the sentence 'Tá mé sásta' (Translation: *I am happy/satisfied*) Note the VSO word order in Irish. A POS tagger for Irish, developed by Dr. Elaine Uí Dhonnchadha, is available online: https://www.cs.tcd.ie/SLP/pos_gle.utf8.html

This process is particularly important in synthesis systems as the pronunciations of words often differ depending on the word class to which the words pertain. The word 'd e i r e a d h', for example, can be pronounced in two different ways depending on the context in which it is uttered. '*Ag deireadh an lae'* [dʲ e rʲ ə] ('at the ***end*** of the day') (noun) vs. *deireadh sé i gcónaí* [dʲ e rʲ ə x] ('he ***used to*** always ***say***') (verb in the imperfect). The synthesiser uses the information generated by the part-of-speech tagger in order to determine the correct pronunciation. The POS tagger for the Irish language has been developed by Elaine Uí Dhonnchadha, TCD (see Uí Dhonnchadha (2010)).

2.6.6. Summary.

TTS synthesis has, to date, had little impact on education generally, or on CALL. The deficiencies in the speech output of earlier systems were such that it was considered unsuitable for all but peripheral activities such as talking dictionaries or the reading of dictation exercises. Its main use tends to be as a reading facility for those with sight difficulties.

In the context of the development of task-based CALL games, the only viable means of producing speech is by way of TTS synthesis if the learner is to be free to interact with the computer in relatively novel ways. This would include, for example, learners

responding to the computer using their own newly-composed utterances - the responses are novel and so the possibility of using pre-recorded speech therefore no longer exists. The computer needs to be able to process the input and to make appropriate responses..The present project uses the newly developed Irish language synthetic speech as part of the interactive CALL platforms being developed. The use of TTS synthesis in a context such as this is rare for any language, and there has been no previous attempt to produce such materials for Irish.

No previous assessment has been done on the quality of the new *ABAIR* synthetic voices. It is crucial that the present study make an assessment of the adequacy of the TTS voices in the context of the interactive CALL platforms which are being developed.

2.7 Summary

Chapter 2 of this thesis presents the literature review for the themes of this thesis as well as the theoretical basis for the development of the three CALL platforms which are described in Chapter 3.2. The platforms are being developed in the context of the teaching of the Irish language. There is, at present, a dearth of CALL materials for Irish and the present project represents one of the few attempts to date to produce task-based interactive language learning games for Irish.

This chapter contains a brief outline of the present position of the Irish language, both as a spoken first language and as a language which is taught to school going children. Irish is a compulsory subject for almost all of the school going population. While there appears to be a positive attitude towards the language in the population at large it is not clear that such attitudes cause a strong motivation for the individual learner to learn the language.

The Irish language has strong Constitutional support and has sound legal backing as an official language both within the State of the Republic of Ireland and as a European

Union language. Yet, it is an endangered language in that it has no monolingual speakers and there are few, if any, domains where Irish is the sole acceptable language. Motivation to learn Irish is mixed. In terms of the school syllabus, no differentiation is made between those for whom Irish is their main home language, those who receive all of their schooling through the medium of Irish and those for whom Irish is just one of their many school subjects.

It is suggested that task-based language learning (TBLL) activities are particularly suitable to a context such as Irish. The task gives authenticity to the language use situation. This authenticity equates somewhat to that of situations where learners have to negotiate meaning and interact with monolingual speakers of the target language. The virtual world of the game becomes a type of monolingual universe where the learner has come to visit or inhabit.

The creation of virtual worlds which are populated by monolingual speakers of minority languages, such as Irish, may confer further advantages in 'glamourising' the language and giving it parity of esteem with major languages which dominate the modern media. This, it is postulated, may in turn increase the motivation of learners.

This chapter briefly reviewed the happy coincidence between the development of CALL programs and prevailing philosophical/pedagogical theory. A pre-1970s behaviourist approach to language teaching was served well by the relatively unsophisticated technology of the time. Since the 1980s the predominant theory of language teaching has been referred to as a 'communicative' or 'interactionist' approach. Such an approach assumes that people acquire language best when they negotiate meaning with competent speakers of the target language. Learners negotiate through information gaps and acquire the language necessary in order to communicate effectively. Such an interactionist approach is the one being adopted in the present thesis. The thesis aims to produce digital

materials with which the learner can interact, thereby simulating a real-life encounter through the medium of the target language.

One branch of the interactionist approach to language has been referred to as 'task-based language learning' (TBLL). This involves the learner using the target language in order to attain some real-world-like objective. Such tasks involve negotiating meaning with an interlocutor, real or virtual, in order to successfully complete some task. Tasks are defined in semantic or pragmatic terms, i.e. the objectives associated with the task are stated in non-linguistic terms such as acquiring directions, solving a puzzle, etc.

The ability of some computer games to stimulate learning has been reviewed. It was noted that there were strong similarities between the general processes involved in computer gaming and the processes involved in TBLL. They both involve setting appropriate goals, providing verbal/non-verbal input, setting out the rules for the task and defining what constitutes success. Game-based CALL programs should share many of the characteristics of general computer games in that they need to be attractive, engaging and at a level of difficulty which is appropriate to the cognitive functioning of the player/learner. This chapter deduces the key features which are desirable for game-based language learning. The proposed features are deduced from the literature pertaining to both computer gaming generally and TBLL.

CALL platforms of the type being proposed need to be highly interactive and able to respond to input in a non-linear fashion. They must be able to handle novel linguistic input and make appropriate responses. It is also important that they incorporate spoken text into their core architecture. Ideally, one would like to integrate a speech recognition feature but this is not technically possible at this juncture. Synthetic speech, on the other hand, offers much stronger possibilities. As the CALL program is capable of making novel utterance responses, it is no longer possible to use the pre-recorded speech of a native speaker and synthetic speech becomes the only type of spoken text possible. The

synthetic speech being used needs to be of a sufficiently high quality in order to give the learner the appropriate linguistic input so that he/she can fully participate in the game. The present project needs to evaluate the newly produced Irish synthetic voices in the context of interactive CALL platforms. Chapter 3.2 describes the development of three platforms which incorporate the features outlined in the present chapter. They are interactive platforms with integrated synthetic speech. Details of the methodologies used to integrate the synthetic speech into the platforms is given and it is noted that since the *ABAIR* synthetic voices are still at a developmental stage the necessary modifications could be made to the synthesis system so as to ensure a high quality speech output for the platforms. The criteria for evaluating the synthetic speech in the context of these platforms are also explored.

Chapter 3: Research Design

3.1 Overview

Chapter 3 describes the development and evaluation stages of this study. It details the methodologies used to evaluate the TTS synthesis from the *ABAIR* project in the context of interactive CALL platforms.

Section 3.2 of the chapter describes in detail the development of each of the three CALL platforms (namely, *Digichaint, Fáilte go TCD* and *Taidhgín*) which are used in this study. Section 3.3 focuses on setting up the evaluation procedures: Section 3.3.1 considers the appropriate criteria to use when evaluating CALL platforms and draws up the specific criteria deemed appropriate for the present project. Section 3.3.2 considers the evaluation of TTS synthesis and proposes the appropriate criteria for evaluating the present TTS synthesis in the context of the CALL platforms which are developed for the project. Section 3.3.3 considers the most appropriate evaluators to use for the project and Section 3.3.4 reviews the techniques which are used for the collection and analysis of the data. Based on the more general literature review of the themes of the present thesis, as presented in Chapter 2, and on the more specific review of literature pertaining to the research design and evaluation detailed in the present chapter, six hypotheses are formulated in Section 3.4. These hypotheses contain the criteria associated with effective CALL platforms and form the basis against which the project is evaluated. Section 3.5 reviews Chapter 3 and gives a summary of the main points.

3.2 The Development of the Three CALL Platforms

The Irish synthetic voices that have been developed up to the time when the present study was conducted (a female Ulster voice and a male Connaught voice) have been used to explore the use of customised platforms, each of which has its own specific salient features. The present project involves experimentation with three individual platforms

which are designed for 16 – 17 year old learners of Irish who are in senior cycle in the Irish secondary school system and preparing for their Leaving Certificate (pre-university) examination. While each has been selected to be included because of the modernity and attractiveness of its graphics which, it is hoped, will add value in terms of its appeal and motivational potential for learners, they are, however, distinct from one another in their structure, content and approaches.

The present stage of development for all three platforms that are part of this thesis is at a proof-of-concept level. The focus of the study is on the suitability of synthetic voices in Irish for use in CALL. It may be argued that the use of digitally-recorded voices would be the most suitable in the case of two of these platforms (*Digichaint* and *Fáilte go TCD*) since they are based on a limited number of pre-determined speech samples in their current forms. However, it should be noted that the present implementations of these platforms should be seen as prototype shells which could, in principle, be exploited by teachers and further developed, filled with different storylines, etc. targeting different aspects of the language to be acquired. To this extent, the true value of using the synthetic speech is that it would permit open-ended exploitation of the platforms without having to commit resources to recording natural voices each time. In the case of the third platform (*Taidhgín*) the learner may make an unpredictable input and this will be read by the synthesiser. This is the only type of situation in which synthetic speech is the only practical response to the needs of the exercise. If the synthetic voices prove successful in the context of the above platforms then the development of games with higher levels of interactivity where the learner can have more direct linguistic input into the game will be possible. In the context of greater interactivity, pre-recorded voices would not be possible because the input from the learners would be unpredictable.

In all three platforms I have tried to stick as closely as possible to the principles of task-based language learning, as set out in Chapter 2.4. The development of such materials can be both expensive and time-consuming and beyond the practical means and ability of an individual research project. Schrooten (2006) describes the difficulties of creating

ideal materials, given the technical and financial limitations within which his project operated, and concluded that the construction of an idealised platform was frequently not within the capacity of small budget projects. For the present purposes, I was particularly lucky in being able to take advantage of the output from two highly specialised projects being conducted within Trinity College, Dublin, namely The Language Trap (Peirce & Wade, 2010) and Metropolis projects (O'Sullivan & Ennis, 2011). It was possible to mould each of these into a platform which was as close to perfection, in terms of fitness-for-purpose, as one could hope to get.

It was possible to incorporate synthetic voices into these platforms giving obvious benefits for learners of Irish. The voices make available native-like speech models which can deal with novel materials (such as novel dialogues composed by learners/teachers) for use in classrooms, or for other learners who have little or no connection with Ireland and do not have easy access to native speaker models. It would also contribute substantially to the debate on the use of synthetic voices for CALL projects which allow for a high level of interactivity between player/learner and computer.

Because the three CALL platforms are quite distinct from one another, conclusions may be drawn in relation to the relative strengths and weaknesses of each. It is intended that this will add to our knowledge and contribute to the future design of games. A detailed description of each of the platforms, in turn, is given in Sections 3.2.1 (Overview of *Digichaint*), 3.2.2 (Overview of *Fáilte go TCD*) and 3.2.3 (Overview of *Taidhgín*). A CD accompanying this thesis gives a sample video of each of the platforms and how they appeared to pupils.

3.2.1 Overview of *Digichaint.*

3.2.1.1 Original platform: The Language Trap.

Digichaint is an adaptation from the 'Language Trap', a language learning game which was developed by the Knowledge and Data Engineering Group, Trinity College, Dublin, by Dr. Neil Peirce and Prof. Vincent Wade (Peirce & Wade, 2010). The 'Language Trap' was an output of the Elektra project, an EU funded European interdisciplinary research project supported by the priority "Information Society Technologies" (IST) within the sixth framework programme. It was an interdisciplinary research project which aimed to promote technology enhanced learning in a variety of disciplines through the use of modern computer games (see www.elektraproject.org). The game was developed as an online casual educational game for the teaching of German language to students who are preparing for the Irish Leaving Certificate. It is a role playing adventure game which has an interactive dialogue system and features synthetic speech in both English and German. In each of the speech encounters the learner is given a menu of possible speech turns in written form. These choices vary in their lexical and syntactic complexity to allow for learners of varying ability levels. When a particular utterance is chosen it is spoken by the synthetic voice. Each utterance is given a star rating according to its perceived level of difficulty. The learner gathers game points according to the appropriateness of his/her selection.

By way of storyline, the learner takes the persona of a newly qualified secret agent who is given a companion agent to guide him/her. They enter a hotel and mingle with German speakers. The learner is given encouragement and advice by way of a friendly companion or hypertext. It has an accompanying dictionary provided by hypertext but the player loses a star by consulting it.

The developers of the 'Language Trap' prioritise its gaming aspect and claim that the language learning experience arises from participating in the game. They used the ALIGN (Adaptive Learning in Games through non-invasion) system to personalise the

experience of moving through the game for each learner. The learner's 'moves' through the game determine the feedback he/she will get and the directions in which he/she will be pointed.

Figure 3.1. A Scene from the 'Language Trap': a screenshot showing the graphics and a selection of the male and female characters in the 'hotel' area.

3.2.1.2 Adaptation: Digichaint.

Digichaint, adapted from the Language Trap, is aimed at 16-17 year old students with specific application to the Irish language Leaving Certificate syllabus in mind. It is designed specifically for the aural and oral components of the syllabus. Students are expected to have had exposure to the three dialects of Irish and all three have approximately equal representation in the aural examination. To date, the Connaught and Ulster dialects have been incorporated into *Digichaint*.

Like the Language Trap, *Digichaint* is also in the form of an interactive guided dialogue that allows students to progress through a virtual world using the original Language Trap graphics of a hotel and its surrounds. Learners are given the task of conversing with

various characters that form part of the game (see Figure 3.1). The dialogue involves the user selecting phrases from a menu of up to four possible options shown on the screen (see Figure 3.2) as a response to questions/statements produced by the game (see Figure 3.3). These are all spoken with synthetic output from the *ABAIR* TTS synthesiser.

The specific scenario developed for the Irish game involves a search by the principal character for the missing half of his/her winning Lotto ticket. He/she had torn the ticket the previous evening, mistakenly thinking that it was not a winner. He/she is quite sure that the missing half of the ticket is in the hotel and someone may possibly have identified that it had some of the winning numbers and realised the error made. The task is to interrogate the characters encountered without revealing one's true purpose in order to avoid the possibility of being double-crossed. The learner has the task of finding the character who holds the other half-ticket by following hints and directions given by the various characters encountered as the game progresses. There is a negotiation involved at the final stage of the game, when the holder of the other half of the winning Lottery ticket has been discovered, so as to get the best deal when it comes to splitting the winnings.

Figure 3.2. A Scene from *Digichaint*: a screenshot showing four utterance options for learners in the 'hotel' scene in the *Digichaint* game.

The game features linguistic adaptivity. "Adaptivity refers to a technological system's ability to dynamically adjust its behavior and settings to an individual user at use-time" (Wang, 2008, p. 19). The language level of the game adapts to the user's language level: as the user chooses more complex structures, the options on offer become more complex, accordingly. Performance feedback and motivational support are provided through a particular companion character in the game who, when requested, will tell the player that his/her selections are excellent, good, poor, etc. Meta-cognitive hints on how well the player is doing appear as thought bubbles linked to the main character. These include phrases such as "Tá ag éirí go maith liom, caithfidh go bhfuilim ar an mbóthar ceart", (*I'm getting on well, I must be on the right track*), "B'fhéidir gur cheart dom triail a bhaint as frásaí níos deacra a úsáid" (*Perhaps I should try to use some more difficult phrases*). The aim of these hints is to draw the learners' attention to the meta-linguistic aspects of the game. They encourage the learners to have the confidence to work towards the more difficult version of the game if they have been performing well up to that point.

The duration of the game is approximately 25 minutes, varying a little depending on the level of difficulty at which one plays the game and one's agility at navigating the screen.

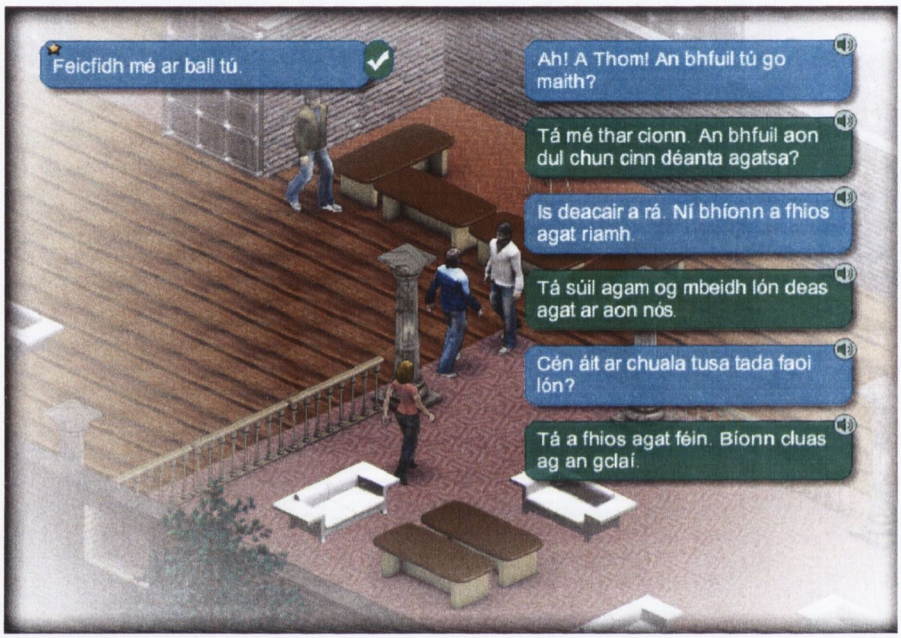


Figure 3.3. A scene from Digichaint illustrating turn-taking: a screenshot illustrating turn-taking in the *Digichaint* game: learner choices are shown in blue, the responses generated by the game are shown in green.

Various forms of personalisation are incorporated into this game. Learners can select their main character by gender or by dialect (male character with Connaught Irish or female character with Ulster Irish). We are currently limited to these two synthetic voices. In order to provide for the more extended cast required for the game, we have added some pitch and speed manipulations in order to differentiate the voices from one another.

Digichaint offers a core design feature which puts the user in control of the direction of the game. The users can determine the other characters with whom they will interact.

Their choice of interlocutors will determine their success in getting through the game and achieving a successful outcome. The linguistic turns are limited to four or five choices from a predefined menu for each encounter. The choices vary both in syntactic complexity and pragmatic suitability for the context. An incorrect choice leads to a loop which directs the player to a set of choices that will set them towards the correct path through the game and does not add to the player's knowledge of the plot. The player must make judgments as to the value of the information which he/she succeeds in eliciting from his/her conversational partner, thereby drawing on his/her cognitive process at a linguistic and at a supralinguistic level.

At the end of the game learners are given feedback on their path through the game, on their star rating and on the words/phrases they looked up in the dictionary (see Figure 3.4). They may transfer this information to their own files for future revision.

Game Summary:

Stars Earned: 256

Stars Lost/Spent: 6

Total number of dialogues: 69

Translations requested: 2

List of translations:

('ar ball': in a while);

('amuigh faoin ngrian': out under the sun);

Figure 3.4. Digichaint feedback: an example of the type of feedback given to the learner after playing the *Digichaint* game.

3.2.1.3 Development of Digichaint.

The Language Trap was chosen because its architectural design allowed for its development into the type of platform which should prove productive for language teaching purposes. Section 3.3 outlines what are considered to be the most desirable features of a CALL game. The following section gives an account of steps necessary for the adaptation of the Language Trap into the readapted game *Digichaint*.

A storyline was first written. This needed to have an Irish context and a legitimate narrative. It had to be written using the appropriate dialects to suit the available synthetic voices. The parts for the female characters had to be written using the Ulster dialect and the parts for the male characters had to be written using the Connaught dialect so that the conversation in the game would sound as natural as possible. The language difficulty level was intended to be such that it would be comprehensible to an Irish Leaving Certificate learner. The options for each conversational turn had different difficulty levels within them and this was intended to give variety to users with different ability levels.

Two versions of the game were written. One for a possible male main character and the other for a possible female. This required differences to the references, for example: in the female main character version phrases such as "maith an bhean" ("*good on you*" said to a female) are used by the other characters whereas if it's the male who is the main character this would be scripted as "maith an fear" ("*good on you*" said to a male). For the purposes of the evaluation, the plot with the female as the main character was used. This was to ensure that all respondents evaluated exactly the same version of the game.

The script was bound by the 'Language Trap' surrounds and architecture. The following features needed to be encompassed:

- scripts needed to be written for 8 characters in total (6 male and 2 female) as well as the main male/female character

- it had to be located in the surrounds of a hotel
- it had to go through 4 different scenes in a predefined order
- it had to meet certain characters in certain scenes in a definite order, as was predetermined by the 'Language Trap'
- incorporated into the script were clues given to complete the task – care had to be taken of the order in which these clues were presented so as to gradually build up a store of information for the main character which allowed him/her to make logical inferences for the completion of the task in hand. One had to be aware of not unduly burdening the cognitive load of the player by having clues too scattered from each other or too subtle

These were the restrictions, but there was also a considerable freedom possible in the adaptation.

- there was no limit on the number of conversational turns which the characters had for any one conversation
- there was no limit on the number of choices offered to the user for each conversational turn
- I was free to plot the steps which a user could take through the game – for example, if they chose an inappropriate utterance they could be sent back to a previous conversational turn and invited to reconsider their options. Due consideration had to be given to the 'flow' of the game so that users would not become frustrated with it and loose interest
- I was free to give between 1 and 5 star ratings to each of the utterances
- a dictionary file was included and there was no limit in the number of entries which it could hold

An introductory text for each of the four scenes was written in English in order to give an overall picture and ensure that every language learner understood the format and plot of game. They are as follows:

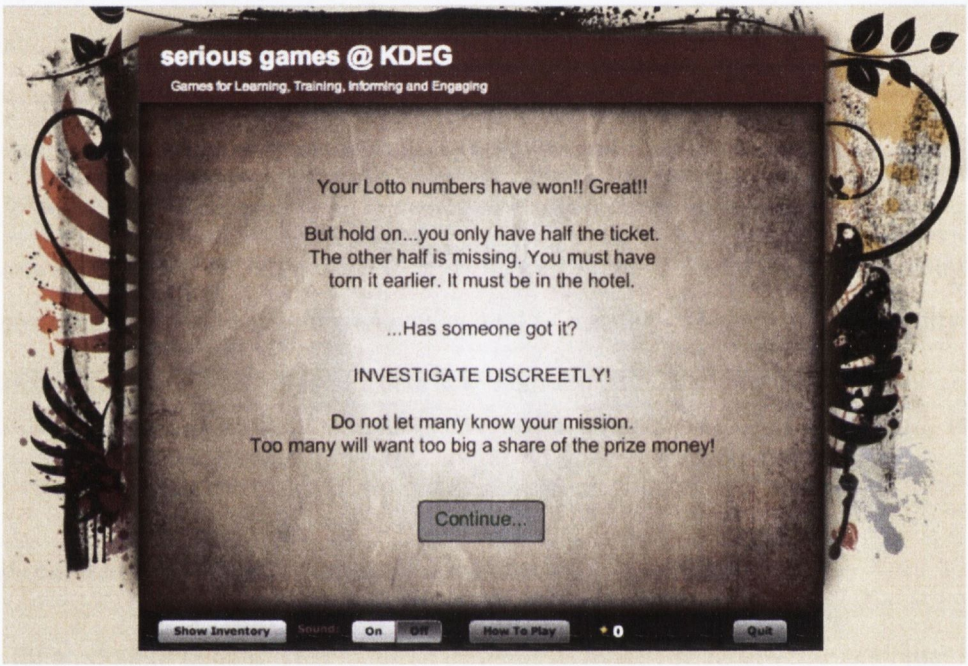


Figure 3.5. Initial Introduction to *Digichaint* – Scene 1 – The Gardens.

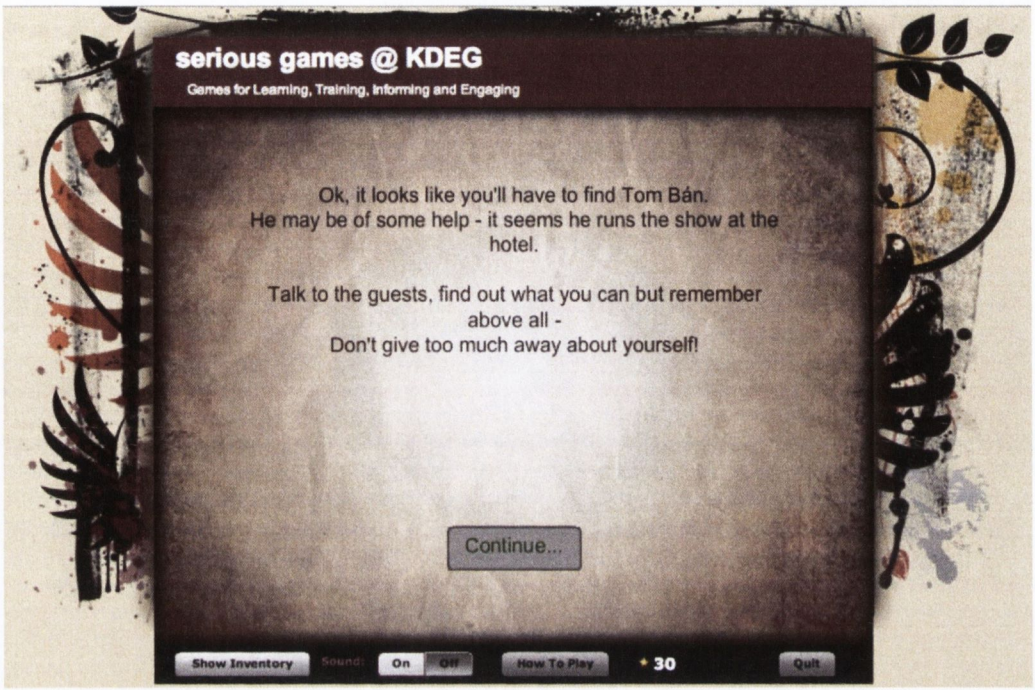


Figure 3.6. Introduction to *Digichaint* – Scene 2 - The Hotel Lobby.

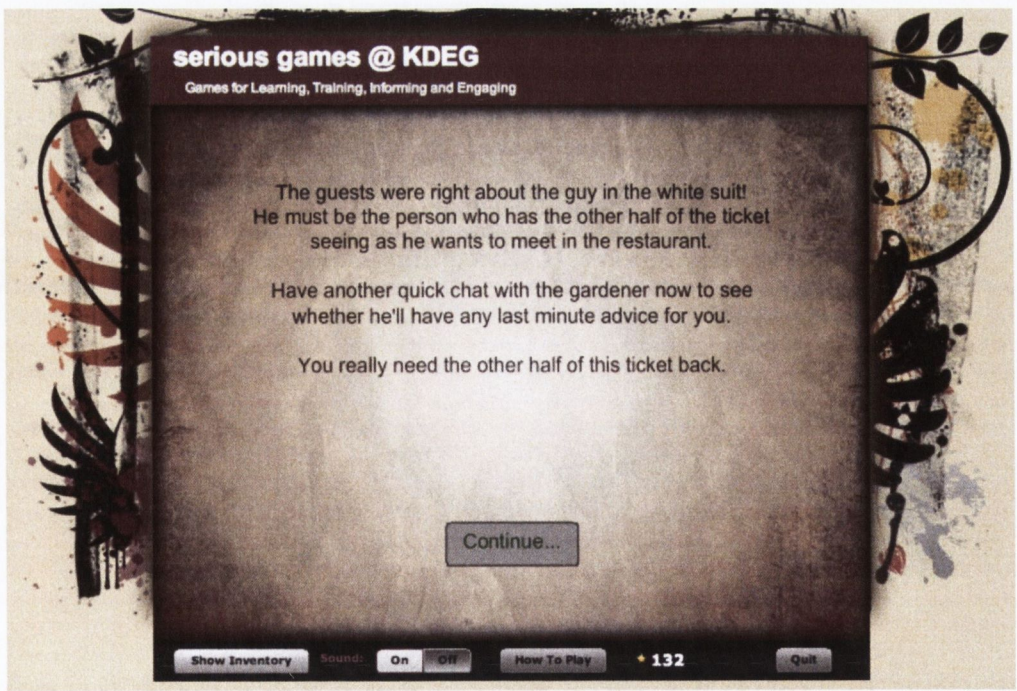


Figure 3.7. Introduction to *Digichaint* – Scene 3 – Outside the Hotel.

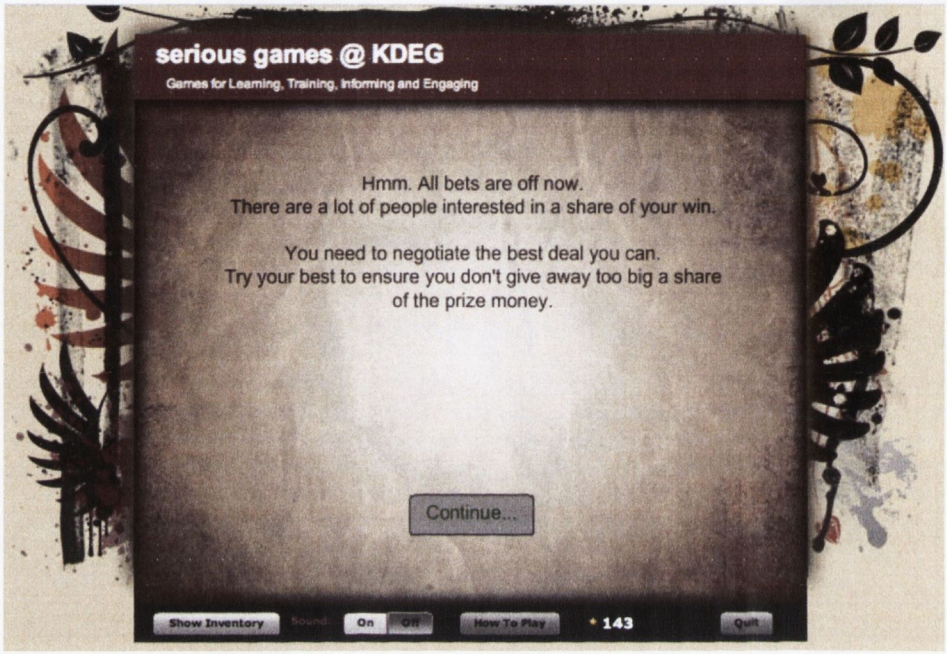
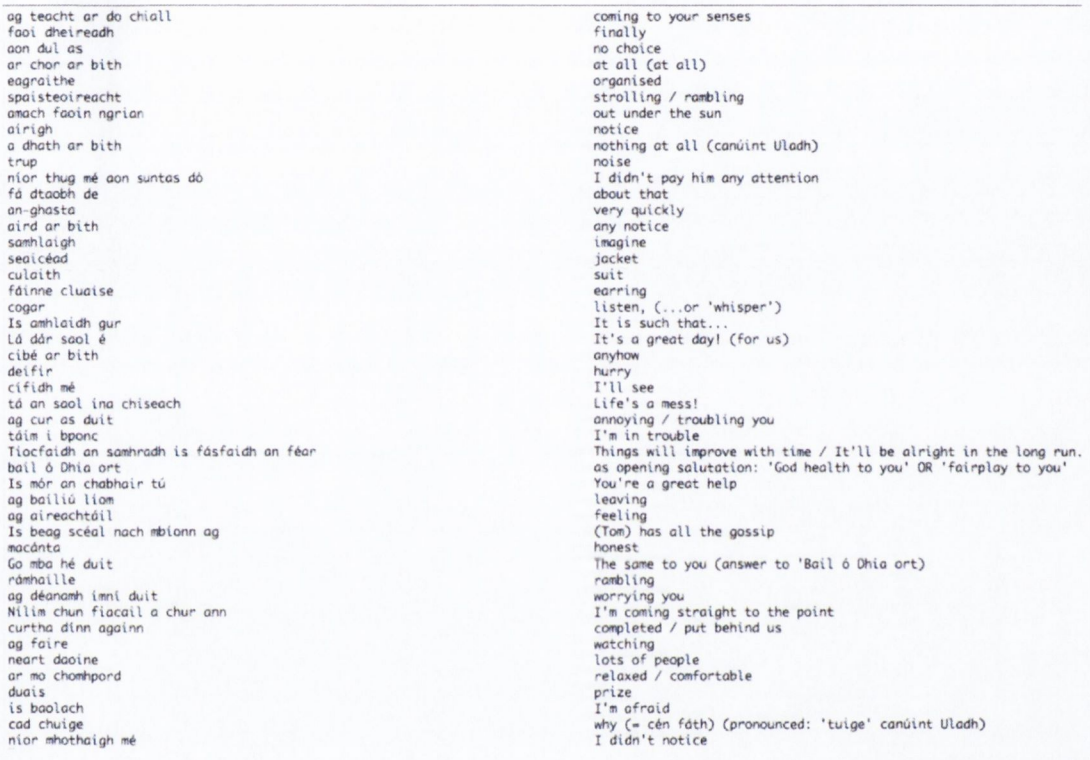


Figure 3.8. Introduction to *Digichaint* – Scene 4 - The Restaurant.

115 dictionary entries, which gave the English translation of particular words and phrases in the context in which they arose in the storyline, were added to the game. Words/phrases, which gave access to the dictionary, appeared underlined in the text presented to the learner. By clicking on an underlined word, the English translation of just that word appeared immediately in the same speech bubble.



ag teacht ar do chiall	coming to your senses
faoi dheireadh	finally
aon dul as	no choice
ar chor ar bith	at all (at all)
eagraithe	organised
spaisteoireacht	strolling / rambling
amach faoin ngrian	out under the sun
airigh	notice
a dhath ar bith	nothing at all (canúint Uladh)
trup	noise
níor thug mé aon suntas dó	I didn't pay him any attention
fá dtaobh de	about that
an-ghasta	very quickly
aird ar bith	any notice
samhlaigh	imagine
seaicéad	jacket
culaith	suit
fáinne cluaise	earring
cogar	listen, (...or 'whisper')
Is amhlaidh gur	It is such that...
Lá dár saol é	It's a great day! (for us)
cibé ar bith	anyhow
deifir	hurry
cífidh mé	I'll see
tá an saol ina chiseach	Life's a mess!
ag cur as duit	annoying / troubling you
táim i bponc	I'm in trouble
Tiocfaidh an samhradh is fásfaidh an féar	Things will improve with time / It'll be alright in the long run.
bail ó Dhia ort	as opening salutation: 'God health to you' OR 'fairplay to you'
Is mór an chabhair tú	You're a great help
ag bailiú liom	leaving
ag aireachtáil	feeling
Is beag scéal nach mbíonn ag	(Tom) has all the gossip
macánta	honest
Go mba hé duit	The same to you (answer to 'Bail ó Dhia ort)
rámhaille	rambling
ag déanamh imní duit	worrying you
Nílim chun fiacail a chur ann	I'm coming straight to the point
curtha dínn againn	completed / put behind us
ag faire	watching
neart daoine	lots of people
ar mo chomhpord	relaxed / comfortable
duais	prize
is baolach	I'm afraid
cad chuige	why (= cén fáth) (pronounced: 'tuige' canúint Uladh)
níor mhothaigh mé	I didn't notice

Figure 3.9. Excerpt from the *Digichaint* Irish-English dictionary file.

Once completed, the narrative for each of the 8 characters had to be written into an xml file. A total of 868 utterances were written for the game. A visual representation of the xml file was available in the form of an Adobe AIR-based dialogue editor developed by Neil Peirce for the 'Language Trap'. The appropriate labelling was done for each utterance by editing either the xml file directly or by using the dialogue editor file as they are interconnected and an update in one automatically transfers to the other. The labelling for each utterance was as follows:

- include star rating
- whether the utterance was considered appropriate/inappropriate in the context
- its standing as an opening turn or a response turn
- the choice of subsequent utterances allowed from that point
- a unique ID for each utterance

```
<dialog character="double_agent" id="Dialog99" x="1507" y="851">
  <statement nextdialog="bye3" id="2" rating="2" correct="true">
    <value>Tá mé cinnte go bhfuil an freastalaí seo sáite sa scéal.</value>
    <response>Is dóigh liom go bhfuil sé ag iarraidh labhairt leat.</response>
  </statement>
  <statement nextdialog="bye3" id="3" rating="2" correct="true">
    <value>Tá an ceart agat ach tá sé ró-dhéanach le tarraingt siar anois.</value>
    <response>Tá sé sin fíor. Caithfidh tú d'ordú a thabhairt don bhfreastalaí.</response>
  </statement>
  <statement nextdialog="bye3" id="4" rating="3" correct="true">
    <value>Tuigim. Ní bhraithim go maith san áit seo.</value>
    <response>Is dóigh liom go bhfuil an freastalaí ag iarraidh labhairt leat.</response>
  </statement>
</dialog>
<dialog character="double_agent" id="Dialog222" x="2130" y="408">
  <statement nextdialog="bye4" id="3" rating="3" correct="true">
    <value>Gabhfaimid beirt go Baile Átha Cliath mar sin.</value>
    <response>Ceathrar againn a bheidh ag dul ann is baolach!</response>
  </statement>
</dialog>
</dialogs>
```

Figure 3.10. Sample of the *Digichaint* game narrative as written in xml.

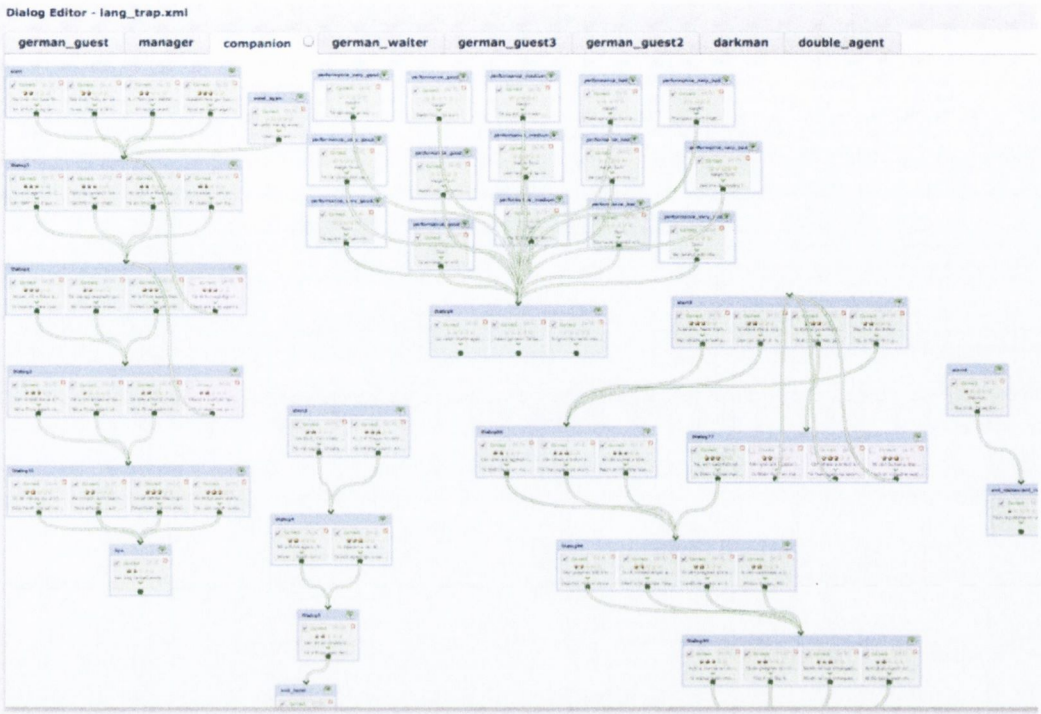


Figure 3.11. Sample of the *Digichaint* game narrative in dialogue editor mode.

Figure 3.11 above gives a visual representation of the possible conversational turns between the main character and another character. They can meet on a possible six occasions and the turn-taking plot for each of the possible meetings is represented above.

An enlargement of one such meeting is represented in Figure 3.12 below. On the top line is the initial choice of four utterances representing four different levels of difficulty. Picking any one of these four leads to the next line where there is another choice of four. On the third line there is one inappropriate utterance which, if chosen, leads one back to Line 2. Similarly, on the fourth line there is one inappropriate utterance which leads back to Line 3. Similar plots were constructed for the exchanges between the main character and each of the 8 characters.

Figure 3.12. Enlargement of a selection from dialogue editor shown in Figure 3.11.

When the narrative was successfully installed it could be uploaded to the game and the text could be viewed instantly.

Creating Soundfiles for Digichaint

Sometimes there is a difference between the written form in Irish and the spoken form. For *Digichaint*, I needed to keep the speech as natural as possible and true to the dialect. In order to do this, variations (indicated in boldface below) on some utterances were created, for example:

- mé f**éi**n (written) -> mé f**hé**in (spoken) [*Translation: myself*]
- **nóiméad** (written) - > **bomaite** (spoken: Ulster dialect) [*Translation: minute*]
- cuar**dach** (written) -> cuar**tú** (spoken: Connaught and Ulster dialects) [*Translation: search*]
- é**igin** -> é**icinnt** (spoken: Connaught dialect) **inteacht** (spoken Ulster dialect) [*Translation: some*]

These were entered into a separate xml file. I now had two separate xml files for the same game – one which dealt with the written text to be presented on screen in the game and the other which was used as input to the synthesiser.

A python script was used to connect the 'spoken' xml file to the *ABAIR* synthesisers. All female characters were directed to the Ulster dialect synthesiser and all males to the Connaught one.

The output from the synthesisers came as 868 .wav files (see Figure 3.13), effectively, incorporating the sound, and 868 separate xml files (see Figure 3.14) which plotted the phonetic transcription of each of the utterances individually. The sound files were then inspected. In the case where concatenation errors/unintelligible output was generated by the systems' letter-to-sound rules I was then able to inspect the relevant xml file and manually adjust the phonemes. This was reprocessed in order to generate a higher quality sound output. The corrections made became upgrades for the synthesisers in that in future when these sounds are entered the synthesiser will use the upgraded version.

Figure 3.13. Soundfiles as generated by the *ABAIR* synthesis system.

```
<?xml version="1.0" ?>
<utterance input_string="T&#xE1;á an praghas sin r&#xF3;ó&#xE1;árd." name="20110813_032322.xml">
<sentence input_string="T&#xE1;á an praghas sin r&#xF3;ó&#xE1;árd." soundfilename="20110813_032322_001">
<token input_string="SILENCE_TOKEN">
<word input_string="SILENCE_TOKEN" trans_source="Utterance.py" trans_output_format="final">
<syllable>
<phoneme symbol="sil"/>
</syllable>
</word>
</token>
<token input_string="T&#xE1;á">
<word input_string="t&#xE1;á" chunkstart="1">
<syllable stress="1">
<phoneme symbol="t"/>
<phoneme symbol="aa"/>
</syllable>
</word>
</token>
<token input_string="an">
<word input_string="an">
<syllable stress="1">
<phoneme symbol="@"/>
<phoneme symbol="nn"/>
</syllable>
</word>
</token>
<token input_string="praghas">
<word input_string="praghas">
<syllable stress="1">
<phoneme symbol="p"/>
<phoneme symbol="r"/>
<phoneme symbol="ai"/>
<phoneme symbol="s"/>
</syllable>
</word>
</token>
<token input_string="sin">
<word input_string="sin">
<syllable stress="1">
<phoneme symbol="sj"/>
<phoneme symbol="i"/>
<phoneme symbol="nj"/>
</syllable>
</word>
</token>
<token input_string="r&#xF3;ó&#xE1;árd">
```

Figure 3.14. xml files as generated by the *ABAIR* synthesis system.

In the final phase of creating the soundfiles for *Digichaint* adjustments were made to the pitch and speed of the original output from the synthesisers so as to make the various characters sound different from each other (effectively, generating 12 characters from the available 2 synthetic voices). This was done using Ableton Live, software designed, essentially for creating, producing and performing music. This system for adjusting the voices was chosen simply as the best available at the time. Adjustments were made to the pitch and speed of the speech output. In the extreme cases, this resulted in rather unnatural voice qualities with, perhaps, a menacing ring. This is the only option currently available although it is not optimal. It is known that changes (especially large changes) in a speaker's pitch entail concomitant changes in voice quality (spectral tile, etc.), and current parallel research in the Phonetics and Speech Lab at Trinity College is directed at elaborating the kinds of changes involved and how they can be incorporated in synthetic speech. Altering the pitch alone in synthesis will impoverish the naturalness of the output (Gobl, Bennett, & Ní Chasaide, 2002). Note, however, that care was taken so that the changes made to pitch/speed did not result in any major loss of intelligibility. Given the fact that some of the characters in the game inevitably had less natural, and sometimes even rather menacing voice qualities, one must expect this to be reflected in the ratings given to the synthesis **quality**, and **attractiveness** in the evaluations.

Current research in the Laboratory is directed at developing the HTS synthesis system in *ABAIR* in such a way as to allow adjustments to voice quality and pitch and it is hoped that this will allow us to 'create' a more natural sounding cast of characters for these games, on the basis of a limited number of 'original' voices. In the present implementation, an effort was made to ensure the adjustments suited the characters, for example, the villain of the plot has his speech adjusted to a considerably lower pitch which gives this voice a menacing and rather eerie ring. The speech for the main characters, i.e. the principal male and female, was left unadjusted and the rest of the cast had to sound significantly different from them. The main characters had the best version of the voice because they had a bigger conversational output than any other character.

3.2.2 Overview of *Fáilte go TCD.*

The *Fáilte go TCD* platform is primarily geared at the development of listening comprehension skills. The platform is based on a virtual reality scene developed by Dr. Cathy Ennis and Professor Carol O'Sullivan as part of the Metropolis Project at the Graphics, Vision and Visualisation Group, Trinity College, Dublin. The Metropolis project was funded by Science Foundation Ireland. For more information see the project website (http://gv2.cs.tcd.ie/metropolis) and O'Sullivan & Ennis (2011).

In this virtual reality scene, the user can navigate around the Front Square of Trinity College, mingling with the crowd, and may eavesdrop on ongoing conversations among groups of students.

Figure 3.15. Virtual conversers in the Metropolis scene (*Fáilte go TCD*).

Graphics for this platform were developed for purposes not related to language learning. They were intended as an emulation of real-life movement by people. The animated characters are created using motion capture data of real conversations between actors originally conducted in English. Motion capture involves recording the movements of the joints of actors: these movements are then transposed to the animated characters in a 3D

application. Since its mainstream introduction in the early 1990s, motion capture has become one of the major tools for 3D character animation (Mitchell, 2012, p. 76).

I have been able to exploit this scenario in order to create a multimedia language learning tool by adding synthetic Irish dialogues to the movements of the characters. The purpose of using this platform was to investigate if a visually rich virtual environment with added synthetic Irish voices would be stimulating for second level learners of Irish. It is intended that the dialogues would be authentic in terms of both their language content and the life-like motions of the characters. This particular platform was selected in order to study the validity of views that the nearness of the graphics or avatars to real-life human interaction, the better. Doughty and Long (2003, p. 58) claim that "new knowledge is better integrated into long-term memory and more easily retrieved if tied to real-world events and activities". "Relevance can also be increased by building realistic characters (Dickey, 2007) or placing the game within familiar environments (Warren, Dondlinger, & Barab, 2008)" (McClarty et al., 2012, p. 20). I tried to have the topics of the conversations as authentic as possible. They refer to themes such as student employment and social activities.

The platform offers learner control in the sense that learners can manipulate the screen so as to zoom into different conversations which are happening simultaneously in the Front Square of Trinity College, Dublin. The full version of the platform has multiple conversing groups, and for *Fáilte go* TCD six short interactions involving different conversing groups were prepared. One two-minute video excerpt was selected to be presented to the learners in the present study. This was done for evaluation purposes so as to ensure that all those evaluating the program were looking at the same graphics and listening to the same language content.

Fáilte go TCD, as it is used for the purposes of this project, offers neither choice of linguistic turns nor choice in the manipulation of the characters, as was integral to

Digichaint. Its strongest features are the clarity and familiarity of its setting, the quality of the animation and the highly developed sense of body movement of the virtual characters.

The educational objective of this particular adaptation is to train listening comprehension skills and to act as a stimulus for classroom discussion in Irish, the target language. The linguistic content of each dialogue is tailored so as to contain particular vocabulary and grammatical constructs and to cover specific conversational topics. These are drawn from the Leaving Certificate (pre-university) examination curriculum, and are pitched at an appropriate level of difficulty. Each dialogue is between three and five minutes in duration, and has been written for a specific group of virtual conversers in the scene.

The multimedia scenario presented in *Fáilte go TCD* is particularly useful for the present purposes in that it acts as a platform on which the TTS synthesis may be readily tested. In principle, it would be, technically, a relatively easy task for teachers or learners to add new dialogues to the groups of conversing characters. One could envisage at a future date that this type of platform could be used in a language teaching situation so that learners or teachers could add new text to complement their material. It could act as a stimulus for learners to use their own creativity in the production of new scenes and its use need not be confined to that of the learning of Irish. In this type of scenario, TTS synthesis would form an essential part of the platform since the dialogues would be unpredictable to the program developer.

This type of activity may be seen as promoting autonomous learning in that it may supplement classroom activities, allow the students to use their computer to learn individually and facilitate their own creativity in producing new dialogues for a virtual world (Abu Shawar & Atwell, 2007b).

3.2.2.1 Development of Fáilte go TCD.

Access to six different excerpts from the Metropolis program were made available to me. These excerpts were composed of groups of three characters in different parts of the Front Square of TCD. The camera begins with an aerial view of the Square and can zoom in to different points where the characters are conversing.

I first decided on student employment and their financial positions as a topic which would be of relevance. I then studied the body movements, the order in which the characters spoke, and the general mood of the excerpt. I constructed dialogues by carefully noting the timepoints and durations of the accentuation or focus in movement of the virtual characters (i.e. where the characters shifted weight from one foot to the other, for example, or stretched their arms in a particular direction). A script was written to fit these time points.

In all, scripts were written and material prepared for 5 videos. Having consulted with a pilot group for their reactions, a 2:07 minute excerpt which showed a conversation between two males and one female was chosen for the evaluation. For each video, text was sent to the *ABAIR* synthesisers in order to generate sound files for the individual sentences in the scripts. As was the case for *Digichaint*, the scripts had to be as true to the dialect as possible. A manipulation of one of the male voices was necessary since there were two male characters in the scene that was chosen for evaluation. This was processed in a similar manner to the voices in *Digichaint* using the Ableton Live software.

The next step in the development was importing both the video footage from the Metropolis environment and the *ABAIR*-generated soundfiles into iMovie, video editing software on the Apple Mac. This software allows you to drag an audio file over the video frame and pinpoint exactly where you would like it to begin playing (see Figure 3.16).

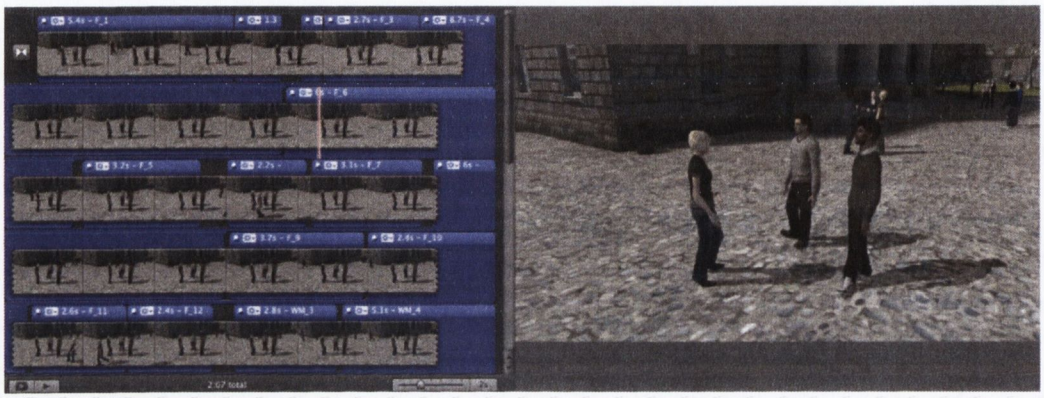


Figure 3.16. Combining audio and video files using iMovie for *Fáilte go TCD.*

It was necessary to divide up some of the sound files. The longer sentences were segmented so as to ensure that focal words in the utterance were synchronised to body movements, such as a wave of the hand or a foot movement, as these tend to be associated with the giving of emphasis to a particular word or phrase. Focal words/phrases were attached to focal movements. This was in an attempt to make the end result seem as natural as possible. See Appendix A2 for full text.

3.2.3 Overview of *Taidhgín.*

Taidhgín is a '*chatbot*' or conversational partner. A detailed review of chatbots is presented in Chapter 2.5.4 above). The name is derived from the phrase in Irish "*Tuigeann Tadhg Taidhgín, ach ní thuigeann Taidhgín tada*" (translation: 'Tadhg understands Taidhgín but Taidhgín understands nothing'). 'Taidhgín' is the diminutive version of Tadhg and the name suggests a subservient role for him in any dialogue interaction.

The most salient feature of this platform is its interactivity. It can process novel utterances from the player and give appropriate responses so that a reasonably realistic discourse can be built between the player and the computer.

Taidhgín was introduced as a means of encouraging linguistic interactivity between the learner and the computer. It was essential to have a TTS system attached to this platform since the learners are free to input self-composed conversational turns. Only a TTS system can deal with this unpredictability in the input from the learner.

The present chatbot, *Taidhgín*, does not claim to have any particularly innovative features in its design. Rather, it is being tested in very limited domains for its suitability as a platform on which the synthetic voices of *ABAIR* may be tested.

3.2.3.1 Development of Taidhgín.

Taidhgín was built using artificial intelligence markup language (AIML), an XML-based open-source programming language which was developed by Dr. Richard Wallace and the Alicebot free software community during 1995 – 2000. *Taidhgín* is hosted and run from Pandorabots, a 'free open-source-based community webservice which enables you to develop and publish chatbots on the web' (Pandorabots, 2013).

Students using this interface could 'chat' with *Taidhgín* by typing natural language questions and comments in a text box and having *Taidhgín* respond with an appropriate answer. *Taidhgín* was designed to hold a conversation with students within a specific domain – the most common introductory topics in the oral examination at Leaving Certificate Irish level. These include: general greeting; name; age; where you live; members of your family; hobbies; favourite holidays. The responses that *Taidhgín* gives contain questions and implicatives that direct the user to continue the conversation. *Taidhgín* also has conditional strategies to direct the user to specific topics when user input could not be recognised.

Taidhgín is represented by a 4-minute mp4 video of a very animated, smartly dressed monkey, typing into a laptop (see Figures 3.17 and 3.18). The video plays on a

continuous loop each time *Taidhgín* produces an utterance and remains frozen while *Taidhgín* is quiet and the student inputs to the system. The animated monkey along with two other suggested avatars were presented to a pilot group of 10 second-level students (the target group for this study) and they were asked to choose their preferred agent. *Taidhgín* was selected as they considered him to be 'entertaining' and 'fun to talk with' and because the video creates the impression of an enjoyable, cooperative dialogue partner.

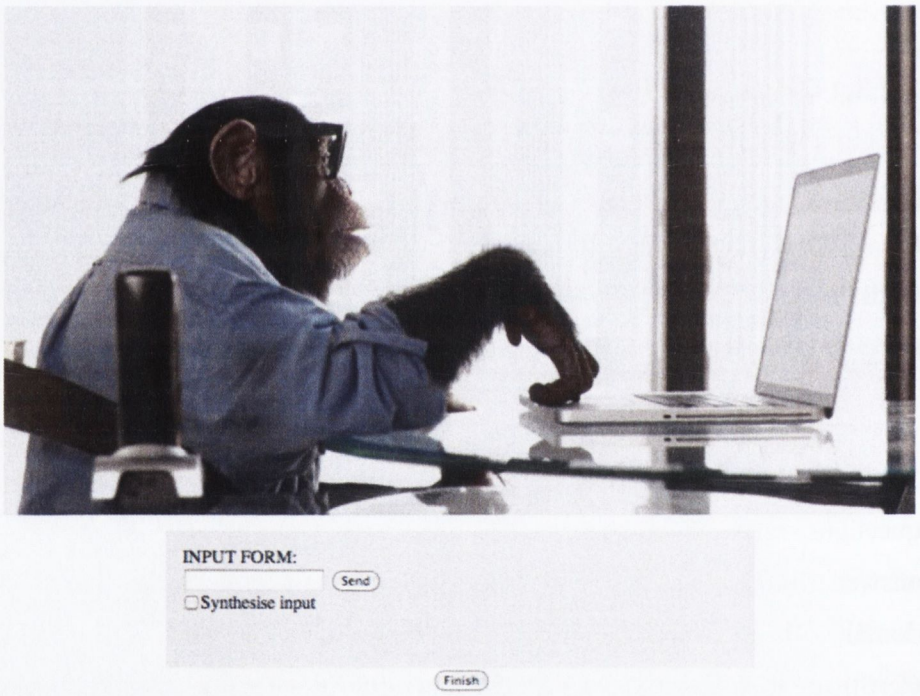


Figure 3.17. Screenshot of Irish chatbot – *Taidhgín.*

Figure 3.18. Taidhgín interface generating appropriate responses.

Pandorabots are available in a format where they are preprogrammed with some set responses to given inputs for a selection of major world languages. In order to use Pandorabots in the present context, *Taidhgín* needed to be developed from new, i.e. the process had to begin with no initial content or knowledge as this was the first time (to my knowledge) an Irish language chatbot had been developed to communicate with users. The aim was to build a repertoire for *Taidhgín* so that he would be able to handle each individual input by the user, find a category with the best matching pattern and return an appropriate answer to the user based on the content/knowledge he was given.

AIML consists of data objects which are made up of topics and categories. Topics are used as the top-level elements and they are given a name attribute and contain a set of

categories related to that topic. Categories are the basic units of knowledge in AIML and consist of at least two further elements – the pattern and template elements. These make up the rules for matching input from the user (the pattern) and linking it to the predetermined answer to output (the template). Figure 3.19 below illustrates the most basic structure whereby the user input 'Hello' (*Dia Duit*) is received and matched by the system with 'Hello, How are you?' (*Dia is Muire duit. Cén chaoi a bhfuil tú?*) being specified as the appropriate output in response.

```
<category>
<pattern> DIA DUIT
</pattern>
<template> Dia is Muire duit, Cén chaoi a bhfuil tú?
</template>
</category>
```

Figure 3.19. Most basic example of 'categories' in AIML.

The first step in developing *Taidhgín* was to create the 'Ultimate Default Category' (see Figure 3.20). This is used when no match is found for the user's input. The strategy used here was to create a list of statements that *Taidhgín* can draw on randomly that don't necessarily follow logically from the statement the user has inputted but directs the conversation back onto something *Taidhgín* can talk about (e.g. '*éist liomsa nóiméad – inis dom faoi do theaghlach*' [translation: *listen to me a minute – tell me about your family!*]). He asks for clarification ('*Ní thuigim i gceart tú?*' [translation: *I don't really understand what you mean?*]) and indicates that he has no response ('*Tá brón orm, tá tuirse orm inniu, ní thuigim an méid atá á rá agat! Abair liom, cén aois tú?*' [translation: *I'm sorry, I'm tired today, I don't understand what you mean! Tell me, what age are you?*]).

```
<aiml version="1.0">

<category>
<pattern>*</pattern>
<template>
<random>
<li>An mar sin é!</li>

<li>Ní thuigim i gceart tú <get name="name/"></get>!</li>

<li>Tá lúb éigin ar lár! Éist, cé mhéad deartháir agus deirfiúr atá agat?</li>

<li>Éist liomsa nóiméad, inis dom faoi do theaghlach?</li>

<li>Má cheapann tusa go bhfuil mise chul dul ag caint faoina leithéid sin! Cé mhéad deartháir agus deirfiúr atá agat?</li>

<li>Ní nós domsa a bheith ag caint faoina leithéid sin! Cé mhéad deartháir agus deirfiúr a dúirt tú atá agat?</li>

<li>Tá brón orm, tá tuirse orm inniu, ní thuigim an méid atá á rá agat! Abair liom, cén aois atá tú?</li>

<li>Gabh i leith nóiméad, inis dom,  <get name="name/"></get>, cá bhfuil tú i do chónaí?<think><set name="topic">CONAITOPIC</set></think></li>

<li>Ná bac leis sin nóiméad! Abair liom arís... cá bhfuil tú i do chónaí?<think><set name="topic">CONAITOPIC</set></think></li>

<li>Ní thuigim anois tú ach ba bhreá liom labhairt faoi do chaitheamh aimsire?
<think><set name="topic">CAITHEAMHTOPIC</set></think></li>

<li>Níor thuigeas é sin anois! Inis dom faoi na laethanta saoire ab fhearr a bhí agat riamh?<think><set name="topic">LAESAOTOPIC</set></think></li>

<li>
An bhfuil aon cheist agatsa le cur ormsa?
<think><set name="topic">HUMANASKSTOPIC</set></think>
</li>


</random>
</template>
</category>
</aiml>
```

Figure 3.20. Screenshot showing 'ultimate default category' in AIML.

At the present stage of development 11 topics consisting of 3,670 categories have been added in order to make *Taidhgín* seem 'intelligent'. The topics are based on the most basic levels of the Leaving Certificate Oral Irish syllabus, as mentioned previously, and include greeting; age; family; place of residence; hobbies and holidays. (See Appendix A3 for a sample path through the *Taidhgín* system as an interaction between a learner and *Taidhgín*).

Students can also ask *Taidhgín* questions – he has a personal profile and an answer for each of the topics about which he converses with students. For example, if a student asks *Taidhgín* where he is situated or where he lives using any of the following phrases (or certain variations on these phrases): '*Cá bhfuil tú i do chónaí? / Cá bhfuil cónaí ort? / Cá bhfuil tú ag cur fút? / Cá bhfuil tú lonnaithe? / Cá bhfuil tú ag maireachtáil? / Cá bhfuil tú? / Cad as duit?*', *Taidhgín* will respond with one of the three following phrases at random:

Version used by *Taidhgín* chatbot	**Translation**
Tá mé lonnaithe i gColáiste na Tríonóide i lár Bhaile Átha Cliath.	*I'm in Trinity College, in the centre of Dublin.*
Táim lonnaithe in oifig thuas staighre i gColáiste na Tríonóide.	*I'm based in an office upstairs in Trinity College.*
Is féidir teacht orm sa tSaotharlann Foghraíochta agus Urlabhra i gColáiste na Tríonóide.	*You'll find me in the Phonetics and Speech Lab in Trinity College.*

If asked whether he has brothers or sisters he answers with one of the following phrases at random:

Version used by *Taidhgín* chatbot	**Translation**
Tá beirt deirfiúr agam, tá siad beirt istigh sa tsaotharlann agus róchúthalach chun caint libh fós! Cuir ceist eile orm!	*I have two sisters - they're both in the lab but are too shy to talk to you for the moment! Ask me another question!*
Tá beirt deirfiúr agam ach níl siad sách forbartha le bheith amuigh sa saol fós! Cuir ceist eile orm!	*I have two sisters, but they aren't ready for the big bad world yet! Ask me another question!*
Ná bí ag caint liom! Tá beirt deirfiúr agam agus tá mé cráite acu! Cuir ceist eile orm!	*Don't talk to me! I have two sisters and they've me tormented! Ask me another question!*

Likewise, if asked whether he ever goes on holidays *Taidhgín* will respond with one of the following phrases at random:

Version used by *Taidhgín* chatbot	**Translation**
** ** An gcreidfeá nach bhfuair mise lá saor riamh i mo shaol? ...Ceist eile! ** **	*Would you believe I've never had a day off in my life! ...Another question!*
** ** Ní bhfuair mise laethanta saoire riamh i mo shaol! ** **	*Never in my life have I had a holiday!*
** ** Ní thugtar laethanta saoire domsa riamh ach bím ag caint le daoine éagsúla ó gach aon áit ar domhan! ...Ceist eile! ** **	*I'm never given holidays but I do talk to people from all over the world! ...Another question!*
** ** Ní gá domsa laethanta saoire a thógáil, ní thagann tuirse ormsa riamh! ...Ceist eile! ** **	*I don't need to take holidays, I never get tired! ...Another question!*

Some miscellaneous features such as 'smart answers' have also been added using language independent AIML compatible tags. One such feature is the <date> tag whereby a student can ask Taidhgín 'Cén t-am é?' (*What time is it?*) and he'll answer using <date format="%I %p"/>, where <date> is set to the nearest hour in the time zone you specify, in this case GMT.

Version used by *Taidhgín* chatbot	**Translation**
Human: Cén t-am é?	***Human:** What time is it?*
***Taidhgín*:** Nílim lánchinnte, timpeall ar 11 AM ar aon nós.	***Taidhgín:** I'm not 100% sure – it's around 11 AM anyway*

Early elements of a possible grammar and spelling correction facility have also been included in the design. The example below illustrates *Taidhgín*'s response when '**dhá dheartháir*' is used instead of '*beirt deartháir*'. '*Dhá*' and '*beirt*' can be translated as

'two' in English but in Irish there are two forms: '*beirt*' is generally used for people only while '*dhá*' refers to anything other than a person. The learner's error and *Taidhgín's* corrected versions are shown in boldface below.

Version used by *Taidhgín* chatbot	**Translation**
Human: Tá **dhá** deartháir agam. ***Taidhgín*:** Ó, tuigim – **beirt** deartháir! Níl aon deirfiúr agat, buachaillí ar fad atá sa teach leat! Agus, an bhfuil na deartháireacha seo níos óige nó níos sine ná tusa?	***Human:*** *I have* ****two*** *brothers.* ***Taidhgín:*** *Oh, I understand –* ***two*** *brothers! You've no sister, all boys in the house! And are these brothers younger or older than you?*

This correction facility can be developed using samples from one's own experience combined with the detailed work by Ó Baoill (1981) and Ó Domhnalláin & Ó Baoill (1978, 1979).

3.2.4 Overall impact of the three platforms.

Together, these three platforms contain suitable criteria for the design of CALL games, as introduced in Chapter 2. *Digichaint* is predominantly task-based with focus on meaning and has a playful spirit. It is strong in allowing for visual and manipulative input by the player and it has a culturally legitimate narrative. *Fáilte go TCD* has a particularly strong visual presentation with exceptional clarity of graphics including the body movement of virtual characters. *Taidhgín* is an attempt at providing interactive dialogue between player and computer. The possibility of the player/learner entering novel utterances in the game is an example of where synthetic speech is absolutely required for the development of CALL games. In the other situations, such as the present versions of *Digichaint* and *Fáilte go TCD*, where there is a limited number of possible responses, all these possibilities could, in principle, be recorded by a human voice. As pointed out earlier,

however, the latter two are intended as examples of material-development platforms which would allow for flexible development of wide ranging materials targeting different aspects of the language and learner groups. The availability of synthetic speech would be essential for such future applications.

When combined, these three platforms have the essential elements necessary for the production of games in an immersive environment which mirrors that of SecondLife (as discussed in Chapter 2.5.3). Each essential element is being tested individually and if they are deemed acceptable by the evaluators, then the unique characteristics from the various platforms may possibly be combined to form a new and more elaborate platform. The chief characteristics of the platform being envisaged are as follows:

- A system which has a culturally legitimate narrative
- A system which allows learners to manipulate the characters and navigate their way through a game which has a realistic plot and clearly defined goals
- A system which requires active problem-solving by the learner requiring linguistic and extralinguistic skills
- Real-world-like graphics and avatars
- The ability to produce a high level of linguistic interaction with the learner
- A system which includes adaptivity to the learners' level of proficiency and offers features such as dictionary look-up and corrective feedback

Proposals for the future development of CALL games, which offer the characteristics outlined above, will be made in the concluding section of this thesis (see Chapter 6) and the centrality of TTS synthesis to such games will be discussed.

3.3 Evaluation Criteria

Section 3.2 has given a detailed description of the development of each of the three platforms. The present section considers the evaluation of the platforms. It first does a theoretical review of the evaluation methodologies and criteria for CALL platforms in general. From these general theoretical considerations, more specific criteria for the evaluation of the present project are proposed.

Section 3.3.1 begins with a theoretical review of the evaluation of CALL materials in order to develop the appropriate criteria with which to evaluate the platforms set out in Section 3.2 above. The specific criteria are identified and a blueprint for evaluation is drawn up. This is followed by a consideration of the evaluation of TTS synthesis (see Section 3.3.2). Since a major focus of the present thesis is an evaluation of the contribution which TTS may make to CALL, the two elements - namely general platform structure, such as graphics, plot, alignment of character movement etc., and the synthetic speech are being given separate consideration. The question of their interdependence is discussed in more detail at the end of Section 3.3.2.1. Sections 3.3.3 and 3.3.4 give a summary of the evaluation methodologies which are relevant to the present study, outlining the details concerning the evaluators and the statistical analyses which are used in the teacher and learner evaluations, the results of which are described in Chapters 4 and 5 respectively.

3.3.1 Towards an evaluation of CALL.

The evaluation of CALL programs has been a matter of some considerable controversy for the past two decades. Most literature on the topic has concerned itself with developing criteria by which language teachers may assess the usefulness of particular CALL programs for the specifics of their own teaching groups and teaching programs (Leakey, 2011). There is a dearth of research published on the evaluation of new emerging programs or on evaluative criteria which is of specific use to program developers.

Evaluation must be seen as an essential aspect at each stage of program development. Reeder et al. (2004) argue for the development of an online technical manual of review and E/valuation techniques. A number of such evaluation techniques have been developed and are available to the classroom teacher (e.g. Hertz (1984)). These are intended to "assess the quality and potential of a software application before it is used with students" (Squires & McDougall, 1996, p. 147). Levy and Stockwell (2006) recognise the centrality of evaluation in CALL design and development. "CALL is about design, development, and evaluation, as well as research and practice conceived around a ready-made product or generic application" (p. 3).

The essentials of any evaluation procedure are to ascertain whether, or to what extent, or in what circumstances, the product which is being evaluated achieves its stated aims and objectives. If the product is a new car which is intended to accelerate from 0 to 100 km per hour in 10 seconds then the evaluation can be a simple matter of time and speed measurement with due consideration of driver skill and road conditions. Though the evaluation of a CALL program is more complex, the essentials remain the same. The evaluation must be directly referable to the aims and objectives of the program, the extent to which it is likely to achieve these aims and objectives, and the circumstances under which it is most likely to be effective. It must define the learning cohort for whom it is intended, the necessary hardware and software for the delivery of the program and the necessary language and computer skill levels of the participating teachers and learners. Garrett (1991) saw only two absolute criteria for evaluating technology-based language materials:

(a) the language context be correct, authentic and appropriate
(b) the program runs as it should without bugs and without crashing

Bailin (1995) was somewhat more specific in his requirement that the functionality of a piece of software be evaluated not just in terms of its internal working but also in the

context in which it is used. At the end of the last century there was little agreement on essential criteria for evaluating CALL material though a number of guides to selecting coursework were beginning to emerge (Healey & Johnson, 1999). "Evaluation is a matter of judging the fitness of something for a particular purpose" (Hutchinson & Waters, 1987, p. 96). Chapelle's (2001) contribution represented a seminal work in devising a theoretical framework for the evaluation of CALL, though a standardised benchmarked system cannot yet be said to have emerged. A more detailed review of Chapelle's work is given in the following Section 3.3.1.1.

3.3.1.1 A framework for evaluating CALL.

Hubbard's (2006) evaluation framework involved three essential elements:

1. A judgment on the appropriateness of the software for a given language learning setting
2. Identifying the ways in which it may be effectively implemented
3. Ascertaining its degree of success

Evaluation may be done by way of judgmental and/or empirical analysis (Chapelle, 2001).

3.3.1.1.1 Judgmental analysis.

Judgmental analysis involves seeking the opinions of the stakeholders in the project, namely, course designers, teachers and learners (Jamieson, Chapelle, & Preiss, 2005). It is also referred to as 'subjective' analysis. It is frequently based on checklists or forms (Levy & Stockwell, 2006), questionnaires or interviews. Checklists may simply require yes/no answers or they may follow along a Likert scale (Likert, 1932), typically 1: Strongly Agree; 2: Agree; 3: Neither Agree nor Disagree; 4: Disagree; 5: Strongly Disagree. A fuller discussion on the merits and demerits of Likert scales is contained in Section 3.3.4.2 below. Questionnaires with open-ended questions are also used but these

may present problems relating to quantifying the data gathered into some kind of meaningful general statement or statements. Quantifying involves judgments by evaluators who may carry their own prejudices and biases to the situation.

Chapelle (2001) developed a list of six issues or questions for the judgmental analysis of CALL. These criteria assume a task-based approach to language teaching/learning.

1. Language Learning Potential: Does it provide sufficient opportunity for beneficial focus on language form and structures?
2. Learner fit: Is the difficulty level of the target language appropriate?
3. Meaning focus: Is the learner's attention directed primarily towards the meanings being conveyed?
4. Authenticity: Is it of interest to learners and does it have application in other non-classroom situations?
5. Impact: Will it provide a positive learning experience? Will learners develop language learning strategies?
6. Practicality: Are the necessary hardware, software and personnel resources available?

Chapelle further tried to develop appropriate tools to extract empirical data relating to each of these aspects of a CALL program.

3.3.1.1.2 Empirical analysis.

Empirical analysis is based on data collected from some observable actions or events and would normally be expected to have some more objective criteria. It is also referred to as 'objective' analysis. Typical of this is the 'pre-treatment test', 'treatment' and 'post-treatment test'. One would normally have an experimental group which would be subjected to the treatment, and a control group which would not. One postulates that any

differences that arise between the groups from the 'pre-treatment' stage to the 'post-treatment' stage are entirely due to the 'treatment' (Barr, Leakey, & Ranchoux, 2005; Bergström, 2007). One tries to exclude extraneous variables which may influence the observable differences but this can prove to be extremely problematic in the case of any life sciences and particularly in the case of human behaviour. Issues such as motivational factors, individual acceptance of computer games, and situational circumstances may be extremely difficult to control in the case of CALL programs. Data from pre-treatment and post-treatment tests provide little information on the learning process and in order to assess whether a piece of software is having a positive effect on learning one must try to systematically observe the students as they use it (Hubbard, 2006). The validity of post-treatment tests needs to be further considered.

3.3.1.1.3 The validity and reliability of testing in the SLA/CALL context.

Test validity refers to the extent to which particular tests measure what they purport to measure. The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, in a joint publication, refer to validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (1999, p. 9). It is considered the most important issue in psychological and educational testing as it affects the interpretation of test results (Popham, 2008). Typically, test validity is referred to in the context of teaching something to someone, and then assessing the extent to which the learner has acquired the given knowledge. One wants to be sure that the information sought by the test is a valid representation of that which was taught. Test reliability must also be considered. Foster (2006) states that one aspect of validity is reliability. "This refers to the consistency of observations, usually whether two (or more) observers, or the same observer on separate occasions, studying the same behaviour, come(s) away with the same data" (p. 86).

The evaluation of CALL programs in respect of their impact on actual language learning is by far the most problematic aspect of the entire evaluation process. In the empirical evaluation section of her framework Chapelle poses the question "What evidence suggests that the learner has acquired the target forms that were focused on during the CALL task?" (Chapelle, 2001, p. 68). Here we must ask what kind of evidence we would need in order to claim that forms A, B and C were acquired as a result of exposure to a particular CALL program. Language acquisition theory suggests extreme caution when dealing with this matter and the simple pre-treatment test, treatment, and post-treatment test must be seen as inadequate.

Gass (2003) discusses the difficulty in attributing cause to conversational production. Making a direct connection between linguistic input and output assumes a behaviourist approach to language acquisition regardless of the quality and sophistication of the input environment. We cannot assume a direct correlation between teacher objectives and the cognitive functioning of the learner. One must take account of the "unconscious interlanguage knowledge" (White, 2003, p. 37) and a half a century of second language acquisition research which points to language being acquired by a cognitive rule-governed process. The nature and working of these rules is still a very controversial issue which researchers have tried to explain based on the theoretical background from which they come, be that constructivist, behaviourist, nativist or socioculturalist or some combinations of these. The term 'rule-governed behaviour' goes back to Skinner (1966) who followed a behaviourist model of language acquisition. The concept of language acquisition by some kind of systematic rule-governed system was also put forward by Skehan (1998) who had a cognitive approach to language acquisition. Skehan claimed that instruction had no effect on the pattern of language learning.

There is a body of research which differentiates between implicit and explicit learning and suggests that certain linguistic structures are more amenable to explicit language instruction while others are more amenable to implicit learning (DeKeyser, 2003).

Second language acquisition (SLA) theory is still a newly emerging scientific field and the SLA process is not very well understood (Gregg, 2003). One may conclude with the generalisations that:

a. SLA belongs in the realm of cognitive processing

b. The linguistic items acquired in the SLA process are largely dependent on the attention given to them by the learner and their importance in a particular communicative context (Al-Hejin, 2005)

c. Item acquisition is not an absolute, in the sense of it having either an absolute present or absolute absent status. One may have an understanding of a new phrase, never heard before, in a particular context because of getting cues from the social or linguistic environment. One may understand a phrase in a restricted way and have very little understanding of its semantic range. Crosswords test the extent of the native speaker's knowledge of the semantic range of various words. Dictionaries frequently give a range of uses for a word the extent of which we, as native speakers, had not realised previously. Logan (1988) investigated the improvement in performance that came from practice and noted the shift to automatisation of language use that came when the learner had frequent exposure to instances of the target language. A distinction may be made between acquisition at the level of comprehension in a particular context and acquisition at a speech production level. One may further distinguish between various levels of appropriateness in the sense that the learner needs to acquire knowledge of the linguistic or social contexts in which an utterance may be used appropriately. An utterance may be grammatically correct but inappropriate in a particular linguistic or social context. Sykes (2008), for example, produced a platform which focused on the teaching/learning of Spanish pragmatics rather than other features such as syntactic or phonetic constructs.

A distinction may be made between evaluation at a macro and at a micro level. At a macro level, one may test whether a particular group of learners have benefited from the use of a particular program over a reasonably protracted period. One may construct metrics to evaluate factors such as the learners' fluency, vocabulary, grammatical structures, accuracy, etc. and conclude that the learners attained whatever level they have reached because of the influence of the programs/courses to which they were exposed. For the reasons outlined above, one cannot equally ascribe a particular linguistic outcome to a micro input such as a short game. In the case of learners who are beyond the very initial stages in acquiring the target language one cannot be sure that they were not exposed to particular forms previously. Indeed, one would need to define very clearly what is to be understood by the term 'has acquired structure X'. X may be acquired to a greater or to a lesser extent, partially acquired or fully acquired.

To return to Chapelle's question "What evidence suggests that the learner has acquired the target forms?" (Chapelle, 2001, p. 68) we should focus more on the word 'suggests'. In the case of games, or short CALL programs, the only real suggested evidence we can call upon is:

a. our observation of the learners' engagement with the program which is about as near as we can come to the kind of empirical evidence referred to by Chapelle (2001) below

b. the learners' reactions to the program/game as collected through questionnaires or checklists. These would appear to be the best achievable data at this time even in the context of Reeder et al.'s (2004, p. 258) warning that "an introspective assessment by learners, for example, can prove to be unreliable, producing contradictory results at times". Handley (2005) warns that "process-oriented evaluations on their own are not adequate - that a learner has engaged in a process that is believed to promote acquisition does not guarantee that they have acquired the intended TL forms" (p. 109).

In summary, one must be aware of assigning any desirable learning by pupils to be the result of input of a particular teaching programme. Evaluations of CALL programmes can only provide tentative data which is open to interpretation by the data analyst.

3.3.1.1.4 Chapelle's model of evaluation.

An outline of Chapelle's model of evaluation is summarised and laid out in tabular form in Table 3.1 below. It is based on the criteria outlined by Chapelle (2001, p. 55). One is struck immediately by its complexity. It is a mixture of subjective and objective data which is obtained by reference to program developers, students and teachers, all of whom are considered stakeholders in the production and use of CALL materials.

Table 3.1

An outline of Chapelle's criteria for evaluating CALL, based on Chapelle (2001, p. 55)

Criteria	**Source of Data**					
	Program Developer		**Student**		**Teacher**	
	Nature of data:		**Nature of data:**		**Nature of data:**	
Language Learning Potential	Objective	Subjective	Objective	Subjective	Objective	Subjective
Learner Fit	Objective	Subjective	Objective	Subjective	Objective	Subjective
Meaning Focus	Objective	Subjective	Objective	Subjective	Objective	Subjective
Authenticity	Objective	Subjective	Objective	Subjective	Objective	Subjective
Impact	Objective	Subjective	Objective	Subjective	Objective	Subjective
Practicality	Objective	Subjective	Objective	Subjective	Objective	Subjective

The model produces a rubric of 36 cells each containing independent data which cannot be quantifiably related to the data from any other cell. To give relative weightings to the values of each of the cells would be problematic and subjective. The amount of data to be collected seems almost prohibitive and questions about its reliability would arise since those being questioned may experience fatigue from the size of the data gathering instruments, and take shortcuts. Chapelle herself recognised these difficulties and stated that the outcome of evaluation "cannot be a categorical decision about effectiveness" (Chapelle, 2001, p. 53).

This leads to a clear demonstration of Scholfield's assertion that research on its own may just end up with *information*, not judgment and action (Scholfield, 2003). The pursuit of the development of absolute criteria for an evaluative assessment of CALL software in general seems misguided. This is not to despair of developing criteria for the evaluation of a specific CALL program and its fitness "for a particular purpose" in a specific situation (Hutchinson & Waters, 1987, p. 96). The Chapelle model offers a framework within which evaluation models for specific programs may be constructed.

Hubbard (2006, p. 6) sets out five general principles which exemplify this approach:

1. CALL evaluation is situation specific;
2. CALL should be evaluated both judgmentally and empirically;
3. CALL evaluation criteria should come from instructed SLA theory and research;
4. the criteria should be applied relative to the purpose of the CALL task; and
5. the central consideration should be language learning potential.

It is within these general parameters that an evaluation framework for the present CALL games is drawn up. This framework is presented in Table 3.2 below.

3.3.1.2 Criteria for evaluating the present CALL platforms.

The present project involves evaluating the usefulness of three distinctive CALL platforms for the promotion of oral and aural Irish language skills among senior cycle pupils in Irish post-primary schools. They are loosely based on the Irish language Leaving Certificate syllabus which has a strong emphasis on oral and aural language skills. The fact that the platforms may have wider application to other learners is of secondary importance for present purposes. Each platform is intended to promote a task-based approach to language learning as set out in Chapter 2.4 of this thesis. It is hoped that platforms such as these may also encourage autonomous learning, as discussed by Little (2003, 2007). By participating in the activities promoted in the various platforms, learners are taking responsibility for their own learning. The evaluation aims, inter alia, to ascertain if, or to what extent, each of the platforms conforms to the criteria for CALL games, as specified in Table 3.2 below. These criteria include the suitability, clarity and general attractiveness of the non-TTS aspects of the platforms as well as the intelligibility, quality and attractiveness of the TTS synthetic speech. The evaluation objectives contain the core components of Hubbard's (2006) model which requires an evaluation of "learner fit" (language difficulty level; language needs and interest level in material) and "teacher fit" (the degree to which it is compatible with teacher expectations and style and its relevance to the syllabus).

Table 3.2 below sets out each of the criteria being used in the present evaluation, the theoretical background on which they are based, and the evaluation methods being employed:

Table 3.2

Criteria and methods of evaluating game-like CALL platforms. This table also includes references to the theoretical background (in Chapter 2 and in Sections of the present chapter) from which the criteria are drawn.

Criteria	Methods of Evaluation	Theoretical Background
1. Games to be task-based with focus on meaning and on use of target language.	Subjective Likert scale questionnaire Open-ended questionnaire Open-ended reflection	Chapter 2.4 - *Task-based Language Learning (TBLL)*
2. Learners' opportunity to develop meta-linguistic awareness. Language difficulty level. Appropriateness of manoeuvring through games.	Subjective/ Objective Questionnaire Tracking of student manipulation of game by electronic tracking/recording	Section 3.3.1.1.1 - (Chapelle, 2001) – see: *1. Language Learning Potential and 5. Impact in particular*
3. Playful spirit of game. Enjoyment level of game.	Subjective/Objective Questionnaire Observation of students playing game	Section 2.5.7 - (Purushotma et al., 2009) - see: *3. All elements of the game should have a playful spirit*
4. Clarity of game plot. Helpfulness of hypertext dictionary.	Subjective/Objective Questionnaire Electronic tracking of student use of hypertext	See Chapter 2.5.2.1 – and see: (Mitgutsch & Alvarado, 2012) 3. Fiction and narrative: *credible narrative involving an engaging story scenario* Section 3.3.1.1.1 - (Chapelle, 2001) – see: 3. Meaning focus: *is the learner's attention directed primarily towards the meaning being conveyed?*
5. Logical structure of game. Appropriate clues to successfully manoeuvre through game.	Subjective Questionnaire	See Chapter 2.5.2.1 – and see: (Mitgutsch & Alvarado, 2012) 3. Fiction and narrative: *a credible narrative involving an*

		engaging story scenario
6. Intensity of engagement with game.	Objective Observation of intensity of student's engagement with game – recording of percentage of students actively engaged in the game at fixed times through the game	Chapter 2.5.2 - (Csíkszentmihályi, 1988; Franciosi, 2011) – see: *"Flow"*
7. Appropriateness of visual and aural material in game.	Subjective Likert scale Questionnaire	See Chapter 2.5.2.1 – and see: (Greitzer, Kuchar, & Huston, 2007; Mitgutsch & Alvarado, 2012) - 5. *Appropriate aesthetics & graphics*
8. Legitimacy of game narrative.	Subjective Likert scale Questionnaire	See Chapter 2.5.2.1 – and see: (Mitgutsch & Alvarado, 2012) - 3. Fiction and narrative: *a credible narrative involving an engaging story scenario*
9. Clarity of screen layout.	Subjective Likert scale Questionnaire	Chapter 2.5.6.2 - (Mayer & Moreno, 2003) – see: *Key principles of CLT*

3.3.2 Towards an evaluation of TTS.

There exists a considerable body of research into the evaluation procedures which may be used for synthetic speech. Itahashi (2000) has proposed what are probably among the most comprehensive guidelines which outline means of evaluating TTS systems. Designed primarily for Japanese speech synthesis, its general applicability would hold across most languages. It proposes an analysis system which begins with a text analysis phase and goes through intelligibility; syllable articulation; word intelligibility; sentence intelligibility and overall quality to include rhythm, speaking rate, continuity, intonation, nearness to human voice and suitability for the users' purpose. Campbell (2007) refers to the many different ways speech synthesis can be evaluated. These include diagnostic or

comparative evaluations, subjective or objective evaluations, modular or global evaluations, task-based or generic evaluations.

The Expert Advisory Group on Language Engineering Standards (EAGLES) produced a TTS assessment taxonomy to explore the many dimensions of synthetic speech. They distinguish between what they call 'black box' and 'glass box' assessments (van Bezooijen & van Heuven, 1997, p. 485). Glass box assessments focus on testing specific components of a speech system. This is laboratory-based and focuses on the testing of specific components of a system, such as the influence of various linguistic environments on the pronunciation of particular phonemes. Clark and Dusterhoff (1999) investigated differences in intonation using a statistical technique known as Root Mean Squared Error. The glass box assessment tries to use objective techniques which compare the synthetic voice to a human voice in the vocalisation of particular utterances (Navas, Hernaez, & Sanchez, 2002). Kang, Kashiwagi, Zou, Ohtsuki, & Kaburagi (2010) applied a specific mathematical theorem, Bayes' Theorem, in order to provide a measure of the closeness of a list of phonemes to those of a native speaker. The glass box assessment is used primarily as a diagnostic tool by system developers in order to improve the quality of the output of the TTS system being developed (Lampert, 2004). The main problem with this methodology is that there is not always a precise match between objective measures and subjective measures. Some objective measures may be over sensitive compared to the human ear. By converse, synthetic speech may be perfect acoustically but regarded as unnatural by the listener (Cryer & Home, 2010). It is widely claimed in recent literature reviews that synthetic speech has developed to a point where 'intelligibility' is no longer a factor (Mayo, Clark, & King, 2011) and research is focusing more on factors such as naturalness, likeability, prosody, its ability to express emotions, its persuasive abilities, etc. (Cryer & Home, 2010). One should not take intelligibility for granted, however, for a number of reasons. Firstly, it may be 'no longer a factor' for the best performing synthesis systems for English or some other major languages, where synthesis has been developed and refined for many years. It cannot be taken for granted in newly developed systems for a new language, as is the case of the Irish synthesis being evaluated here.

Furthermore, intelligibility should not be seen as an "all or nothing" concept. One may have various degrees of intelligibility and 'ease of intelligibility' (see Section 3.3.2.2 below). It may be entirely unrealistic to expect to develop some objective method which would have a high predictability when postulating on the acceptability of a TTS system to some future hearer or groups of hearers.

Mayo, Clark and King (2005) studied the acoustic features to which users pay most attention. They concluded that people make judgments about naturalness along at least two dimensions: the appropriateness of the prosody and the appropriateness of the units used in the synthetic segments. Podsiadlo (2007), in a much more large scale study, similarly concluded that two things most detrimental to perceived naturalness of the voice are bad joins and "roboticness". Delogu, Conte and Sementina (1998) found that the recognition of a message will be hampered if the sequence of words lacks prosodic continuity. It should be noted that these particular synthesis deficits are linked to different types of synthesis methods that are used. For example, 'bad joins' arise in concatenative unit-selection synthesis where (typically) small segments of a corpus are extracted and joined for synthesis. Roboticness can be an issue with some of the current HTS synthesis systems, and it was also a problem for some of the much earlier formant-based synthesis systems. Alternatively, Andersson, Yamagishi, and Clark (2012) concluded that speaking style is strongly related to peoples' judgments of 'naturalness'. Synthetic speech should have the appropriate discourse features for the context, such as hesitations, agreements, etc. For a reminder of two of the main current synthesis methods, see Chapter 2.6.3. All in all, it is worth remembering that, as synthesis technology is rapidly developing, the issues raised in any one review may quite rapidly cease to be relevant.

A comprehensive review of standardised tests of synthetic speech has been compiled by Lemmetty (1999). These include diagnostic rhyme test (DRT); modified rhyme test (MRT); standard segmental test; cluster identification test (CLID); phonetically balanced word list (PB); nonsense words and vowel consonant transitions. It is important to remember that these features pertain to the intelligibility of the speech, and especially the

success of a system in producing adequate segmental output. As mentioned, intelligibility and naturalness are to a large extent determined by the type (and specific quality) of the synthesis system used, e.g. unit-selection, HTS, or even formant synthesis. As systems evolve and synthesis systems dominant at a point in time may differ, so too will the evaluation criteria and features. It is not proposed to further review these tests in the present context as they do not form the main focus of the type of evaluation which is being considered.

The black box assessment concentrates on the functionality of the overall system and its fitness for use in specific situations such as the telephone answering systems used in banking (Lévy & Hornstein, 1994). This is typically done by way of human evaluation of the synthetic speech. By its nature, evaluation will be subjective and, therefore, a reasonably large number of evaluators will be required in order to control for individual differences between listeners.

Viswanathan and Viswanathan (2005) proposed evaluation criteria covering seven characteristics of the synthetic speech:

1. Overall sound quality
2. Listening effort
3. Ease of comprehension
4. Articulation
5. Clarity of pronunciation
6. Speaking rate
7. Pleasantness

They further argued that the synthetic sentences played to listeners must mimic the application domain to which the TTS system is likely to be used. They were referring here to sentence length and complexity in particular. Hashimoto, Yamagishi, Byrne, King and Tokuda (2011) accepted somewhat similar criteria in testing synthetic speech for naturalness, adequacy and fluency.

3.3.2.1 The evaluation of TTS synthesis in a CALL context.

Handley (2005) noted that the requirements that CALL applications impose on TTS synthesis had not been investigated in any significant way. She found that only one evaluation of the adequacy of TTS synthesis for use in CALL had been conducted and this study related back to Stratil et al. (1987). One must also note not only the fact that synthesis today is very different from what it was in 1987, but also the restricted use to which the synthetic speech was being put. It was merely used for the presentation of grammar exercises in Spanish.

Kang, Kashiwagi, Treviranus and Kaburagi (2008) found no existing research studying whether foreign language learners can accept a synthetic voice or how they perceive or comprehend it. They conducted a study amongst seventy Japanese College Freshmen who majored in business or engineering and simply asked the students if the voices they heard were easy to understand and acceptable. The materials being tested were in two categories. The first category consisted of individual words and the second category consisted of unconnected sentences and involved the students' preferences for natural or synthetic speech. Their overall results showed a slight preference among the subjects for the synthetic voice. They concluded that learners are not as sensitive to differences in naturalness between natural voices and synthetic voices as native speakers are when synthetic voices are easy to understand. It seems that naturalness of speech is not as important a factor for learners as is intelligibility. They rely on the proposition forwarded by Stern, Mullennix and Yaroslavsky (2006) for an explanation of their findings. The latter postulated that synthetic voices are perceived more positively when they are presented as emanating from a computer rather than being the voice of a human. Natural

voices are preferred when they are seen as pertaining to a human. Learners' judgments of the clarity or naturalness of synthetic voices must, quite clearly, be related to the stage of development at which they are. If the synthetic speech being presented to them is of a standard which is higher than that of their comprehension ability, then their opinion on the clarity or intelligibility of the voices will lack reliability. One cannot be sure whether it is their inability to recognise particular words or phrases (because they have never encountered them before) or whether it is the lack of clarity of the synthetic speech that is in question. It would seem that learners can only assess synthetic voices when they are dealing with spoken material which is at or below their language comprehension level. In this context, it is being hypothesised that those with a lower language comprehension ability in Irish will give a lower rating to the quality of the synthetic speech, because, potentially, they do not understand it.

Pellegrini, Costa and Trancoso (2012) report a contrastive study about the use of TTS synthesis instead of pre-recorded utterances in a dictation exercise. Students transcribed 20 sentences spoken with a synthetic voice and 20 sentences spoken with a natural voice. The synthetic utterances were more correctly transcribed than were the human voices. The authors report that the synthetic speech was 15% slower than the natural speech and suggest that this may account for the better performance in the transcription of synthetic speech. This work suggests that learners may not experience particular difficulties with synthetic speech and it offers some optimism for its use as part of a CALL application. Another possible interpretation of the results may be that the learners, having less experience of the target language, may have a lower emotional connection to its prosodic features than that of a native speaker.

Handley (2005) found that different roles for synthetic speech in CALL programs place different demands on the quality of the output. She was referring to its role as a reading machine (RM), as a pronunciation model (PM) and as a conversation partner (CP). She enquired into the adequacy and acceptability of the synthetic speech output at phonetic and prosodic levels and the participants in her enquiry were teachers of French. She

found that ratings of the quality of synthetic voices varied significantly in respect of: precision of phonemes; appropriateness of prosody; naturalness of phonemes and naturalness of prosody, depending on the use to which they were being put. She distinguished four different roles for the synthetic speech:

(a) a reading machine concentrating on the phonetic accuracy of individual words

(b) a pronunciation machine to be used as a model for phoneme pronunciation

(c) a pronunciation machine where the prosodic quality of the output was considered

(d) output which was used as a conversation partner

She used a number of different CALL programs that incorporated TTS synthesis. There were some variations between the acceptability of different programs. It is worth noting that while the different roles given to the synthetic speech caused variations in the ratings given by the teachers to the quality of the synthetic speech, and while many of these variations were statistically significant, they were not dramatically different from each other.

3.3.2.1.1 The influence of context on TTS evaluations.

Campbell (2007) referred to the interdependence of the various components necessary to build a synthetic voice such that a weakness in one component can have unpredictable knock-on effects on other components. This same interdependence between voice and context may be occurring at a supralinguistic level, i.e. listeners' opinions of the synthetic voices may be dependent on their overall view of the platform into which the synthetic speech is integrated. The acceptability of the speech may be dependent on the acceptability of the platform or vice versa. There is also some evidence to suggest that one's prior experience of synthetic voices may have an influence on one's attitude towards a synthetic voice in a particular context (Cryer & Home, 2009).

3.3.2.2 Criteria for evaluating the Irish TTS synthesis in the present CALL platforms

There have been no previous evaluations of a TTS synthesis system applying to CALL-based game-like activities. What evaluations there are in the CALL area, were done in the context of the synthetic speech having some kind of tutorial role such as a model for teaching language comprehension, a pronunciation model or a grammar tutor. The present study aims to fill that gap by studying the acceptability of synthetic speech in the context of language games or game-like multimedia tools.

The present study is confining itself to the application of the overall system to the three particular CALL platforms as set out in Section 3.2. This amounts to an assessment of the usefulness of the output from the *ABAIR* system as an intrinsic part of these CALL applications. The TTS synthesised speech for Irish (*ABAIR*) is still at a developmental stage and has had no formal evaluation process to date. Such a formal assessment needs to be developed as part of the overall *ABAIR* project. Since evaluations on the use of any synthetic speech for CALL purposes are scant at present, it is hoped that the present study will be of assistance in exploring the contribution which TTS systems can make to the further development of CALL.

In considering the criteria which may be used for the evaluation of Irish TTS it was firstly decided that the 'black box' type assessment of the functionality of the speech was the most appropriate approach. In this context, the first criterion of assessment must be intelligibility. Three main criteria have been selected for the evaluation of TTS in the present study.

(1) Intelligibility: while this factor is now generally taken for granted in the evaluation of TTS synthesis for some of the major world languages (Andersson et al., 2012; Mayo et al., 2011), one cannot do so in the present context because no previous evaluation has been done on the *ABAIR* synthetic speech. Intelligibility is obviously the most important

of the criteria for a CALL game with speech output because, without it, one cannot proceed with the exercise. Intelligibility may be gleaned from the social or linguistic environment in which a phoneme or syllable occurs and may not be an accurate measurement of its acoustic closeness to the voice of a native speaker. It may also be influenced by the physiological ability and psychological orientation of the listener (attention, motivation, etc.) as well as his/her previous knowledge of the language and socio-linguistic competence (Antons et al., 2013; Voiers, 1977).

(2) Quality: the second of the criteria chosen is 'overall sound quality' as used by Viswanathan and Viswanathan (2005). This is a generalised, non-technical term which is used to assess the listeners' overall view of the voice. It has a suggestion of aesthetic evaluation (D.-Y. Huang, 2011; Mannell, 1984). The assessors are not expected to have the technical or linguistic knowledge necessary to make specific judgments. 'Quality' is left as a term to get the listeners' generalised view on the overall standard of the synthetic speech. A large proportion of the evaluators in the present study are second level pupils and one can only expect responses of a generalised nature. Overall 'quality' is important as one of the criteria for judging the TTS synthesis in the present project as there is a suggestion in the research literature that users' acceptance of synthetic speech may depend on the context in which it is being used (Cryer & Home, 2008). In terms of the present project this is of particular significance since game playing is regarded as a 'holistic experience', as mentioned in Chapter 2.5.6, and therefore learners' general impression of the quality of the TTS synthesis will have a bearing on their opinions on the quality of all the other various background features of the game, and vice versa.

(3) Attractiveness: the third of the criteria chosen for the present evaluation is 'attractiveness'. This refers to the extent to which the listener is not put off by the voice. There is some evidence that listeners generally prefer human voices to synthetic ones (Francis & Nusbaum, 1999). It has been claimed that computer users may assign a personality to the voice and may be strongly influenced by its characteristics (Nass & Lee, 2001). A very negative attitude towards the attractiveness of the voice would make it

less useful as an element in a CALL platform as learners may be less inclined to persist in the activity being promoted in the exercise.

In essence, one is attempting to measure the intelligibility and acceptability of the synthetic speech. Cryer and Home (2010) describe the difference between intelligibility measures and acceptability measures as 'performance measures' and 'opinion measures'. Performance measures give information on whether the voice is sufficiently accurate to be understood while opinion measures give a measure of the acceptability of the speech for specific listeners in specific contexts.

As stated in Section 3.3.2 above, intelligibility cannot be seen as an "all or nothing concept". The concept of 'ease of intelligibility' is being introduced in order to capture the ease with which a particular piece of speech may be understood by the listener. The term 'intelligibility' here is not being confined to a recognition of individual phonemes or words in isolation as used by a number of investigators towards the beginning of the century such as Koul (2003). Rather, it is used as the listener's ability to recognise (and orthographically transcribe) a set of sentences but not requiring any higher level cognitive functioning such as constructing a coherent mental representation of the information contained in the text and relating it to any pre-existing knowledge which the listener may have (Kintsch & van Dijk, 1978). One may, for example, transcribe a spoken utterance with complete ease if it is spoken with clear diction in a manner that is familiar to the transcriber. At this developmental stage of the production of an Irish synthetic voice, it is unlikely that the synthetic speech has reached a level in its development where it is as readily and easily intelligible to the listener as a natural voice. In other words, it may be more cognitively demanding on the listener to decipher than would be the case for natural speech. I am proposing the concept of 'clarity', which brings together both 'intelligibility' and 'ease of intelligibility' measures (or 'performance' and 'opinion' measures), to cover this phenomenon. Details of a *Clarity Index* developed for the present study are outlined in Section 4.3.2.2.2.

3.3.3 Evaluation and evaluators.

The present study poses particular evaluation problems which in turn influence the categories of evaluators used. The evaluation problems are based on the fact that one is simultaneously dealing with three unknowns:

(a) the value of the previously untested CALL platforms being developed

(b) the intelligibility, quality and attractiveness of the *ABAIR* synthetic voices (there has been no previous formal evaluation of the *ABAIR* TTS system)

(c) the contribution which the *ABAIR* output can make to the development of the CALL games

One is attempting to evaluate the non-linguistic (i.e. non-TTS) aspects of the CALL platforms (graphics, plot, alignment of character movements, etc.) and to simultaneously evaluate the *ABAIR* synthetic voices (TTS output).

It was decided to use Likert scale-based questionnaires as the main method of eliciting the evaluators' opinions in this study. This is based on Chapelle's (2001) discussion on the judgmental analysis of CALL (see Section 3.3.1.1.1) and the methods she proposes in order to ascertain whether or not the learner has acquired the target forms (see Section 3.3.1.1.3) – i.e. observing learners' engagement and judging the learners' reactions to the program/game as collected through questionnaires. Likert scales are discussed further in Section 3.3.4.2. Questionnaires were formulated in order to ask for the opinions of respondents in relation to the various factors being examined. Open-ended dialogue boxes were also included as part of both the teacher and learner questionnaires in order to give evaluators the opportunity to give either positive or negative feedback on the CALL platforms and on the integrated synthetic voices. The responses elicited from these questionnaires form the subjective data in the present study and both teachers and learners provide these data. Objective data are elicited from teachers only and the development and results pertaining to these data are described in detail in Chapters 4.3

and 4.9. More detail on the questionnaires and on the theoretical background on which the individual questionnaire items are based is given in Chapters 4.4 (Instruments: Teachers) and 5.4 (Instruments: Learners).

In evaluating the non-linguistic aspects of the CALL platforms it was decided that the appropriate evaluators would be pupils of a similar age to the target population for the platforms – i.e. 16–17 year olds.

Deciding on appropriate evaluators of the synthetic speech was more problematic. There is evidence to suggest that one's proficiency in the target language would be a significant factor in one's judgment of the synthetic voices (Kang et al., 2008). Lecumberri, Cooke and Cutler (2010) made a comparison between native and non-native speaker performance in listening comprehension when the speech was presented against a background of noise. They aimed to clarify the role of prior linguistic experience in speech perception. They concluded that even true bilingual listeners never quite reached the ability of monolinguals in the presence of noise. Synthetic speech is quite clearly less than an idealised form of speech and lack of naturalness may be somewhat similar to the distraction caused by noise. One would expect that prior linguistic knowledge would be a significant factor in determining one's ease of comprehension of the synthetic speech.

Based on the assumption that one's proficiency in the target language is a significant factor in influencing one's evaluation of TTS output, it was decided that different categories of evaluators should be included.

The first category included second level pupils for whom the platforms are developed. As discussed in Chapter 2.2 above, second level learners of Irish come from a variety of socio-linguistic backgrounds. It was decided that participants from each of the three school types should be included in the study – i.e. those from English-medium schools, Gaelscoileanna (Irish-medium schools outside the Irish-speaking Gaeltacht areas) and

Gaeltacht schools. A range of language comprehension abilities would therefore be represented. This is important since it has been argued that high comprehension ability in Irish is needed in order to make judgments on the synthetic speech. The language level of the games has been designed for those from English-medium schools, however, and therefore a significant sample from that school type is also required. A more detailed description of the learner evaluators is given in Chapter 5.2.

In addition to learner evaluators, it was also decided that teachers of Irish should be used to evaluate the synthetic speech. The assumption being made here is that these teachers have a high language competency level and that any defects which they find in the speech are genuine faults with the TTS synthesis rather than lack of comprehension ability by the listener. It was further decided that more objective evaluations of the text could be done using the teacher sample, in that they could reasonably be asked to orthographically transcribe relevant sentences. Details of this orthographic transcription task are outlined in Chapter 4.3.2.2.1. It is in the context of these orthographic transcriptions and teachers' ratings of the 'ease' with which they could carry out the transcriptions that a formula for producing a *Clarity Index* is presented (see Chapter 4.3.2.2.2). A full description of the teacher participants is given in Table 4.4, Chapter 4.9.2.

The following is a summary of the four categories of evaluators:

(a) teachers of Irish who had a high proficiency level in the language
(b) learners of Irish from second level schools from 4th and 5th years (16-17 year olds)

The learners were subdivided into three groups:

1. pupils from Gaeltacht schools, some of whom had Irish as their main home language. These pupils were all from rural backgrounds.

2. pupils attending Gaelscoileanna, Irish-medium schools outside Gaeltacht areas. These pupils were mainly from urban areas and had English as their first language.
3. pupils from English-medium schools who studied Irish as a school subject only. There was a mixture of pupils from urban and rural backgrounds here.

3.3.4 Statistical analysis.

3.3.4.1 Overview.

The statistical analysis for this study is designed to create data which have been abstracted from both subjective and objective measures. In this section, a brief review is given of the statistical techniques used and the assumptions which underline them.

The objective measures are dealt with exclusively in the teacher evaluation, described in Chapter 4. The subjective measures (in both the teacher and learner evaluations, Chapters 4 and 5 respectively) are elicited by way of questionnaire items which ask for the opinions of respondents in relation to the various factors being examined. Respondents were asked to react to questionnaire items by marking their opinions on a Likert scale of 1 to 5. Likert scales typically range from 'strongly agree' to 'strongly disagree' with the proposition put forward (see Section 3.3.4.2). The data abstracted from these responses are further analysed using non-parametric analysis techniques in order to establish the relationship between certain background factors pertaining to the respondents and the opinions they expressed on the Likert scale-based questionnaire items (see Section 3.3.4.3). The statistical technique used to establish the relationship between these responses and background factors pertaining to the respondents is the Kruskal-Wallis test. The Kruskal-Wallis test is described below under Section 3.3.4.4.

The Spearman's Rank Correlation Coefficient is used in order to test the relationship between responses given to the totality of the post-game questionnaire items. This tests

the proposition that respondents experience the platforms as a holistic experience and that their reaction to any one aspect of the platform will be related to their reaction to other aspects of the same platform - for example, if they have a low opinion on the graphics, plot, etc. then they are likely to have a low opinion on the quality of the synthetic speech. This will test the premise that one's opinion on the quality of synthetic speech is not necessarily consistent, but rather, varies with the context in which the speech is used. This test is also described below under Section 3.3.4.4.

3.3.4.2 Likert scale.

It was decided that primary data should be elicited from the school pupils for this study by way of 5-point Likert scale items. Likert scale items sometimes contain 7 or 9 point scales. The present study followed the Viswanathan and Viswanathan (2005) recommendation and decided on a 5-point scale with each of the 5 labels representing more appropriate or meaningful statements such as 'Strongly Disagree; Disagree; Neither Agree nor Disagree; Agree; Strongly Agree'. This produces a symmetry of items about a middle category. Longer scales may require more arbitrary statements and the relationship of one statement to another may be less clearly defined. A number of factors such as social desirability may pose a threat to the validity of the data obtained (McLeod, 2008). Individuals may want to give what they perceive as the desired responses or responses which would not cast themselves in a bad light. McLeod (2008) suggests, however, that this desirability factor may be reduced when respondents are offered anonymity. In order to offer this increased anonymity in the present study, usernames ('user101'; 'user102'; etc.) were randomly assigned to each of the respondents, removing the need to use their actual names. This practice could be expected to lend high validity to learners' subjective responses. Test validity also needs to be taken into consideration here as it has been claimed that the scores on Likert scales may lack validity because of respondents wishing to give the 'right' answer in the interests of being polite (Worcester & Burns, 1975). One needs to be cautious in terms of the procedure involved in administering the questionnaires (described in Chapter 4.5 (teachers) and Chapter 5.5 (learners)) and to limit potential pitfalls and then to survey the findings (see Chapter 4.9.4

(teachers) and Chapter 5.9.3 (learners)) to try to ensure insofar as possible that teachers/learners have given honest answers to the questionnaire items.

3.3.4.3 Non-parametric data.

The appropriate method of statistical analysis of data gathered from Likert scales is a subject of some considerable debate. This boils down to whether the data can be regarded as 'interval-level data' or 'categorical data'. The former refers to a measurement continuum, for example, length or temperature. One can then derive a mean, e.g. the mean length of 3 pieces measuring 5 meters, 9 meters and 10 meters is 8 meters. By contrast, 'categorical', or 'ordinal' data, have a value assigned to them, which is arbitrary and has no unique mathematical property. It is argued, however, that intervals in well constructed Likert scale items may approximate to what is referred to as an interval-level measure. Cohen, Manion and Morrison (2011), however, contend that it is not appropriate to infer that the intensity of feeling between 'Strongly Disagree' and 'Disagree' is equal to the intensity of feeling between other categories such as 'Disagree' and 'Neither Agree nor Disagree' (p. 387). The concept of 'mean' becomes somewhat meaningless in that "the average of 'fair' and 'good' is not 'fair and a half'" (Jamieson, 2004, p. 1218). Pallant (2010) states that non-parametric techniques of statistical analysis are best used when dealing with ordinal data. She contends that a mean of 3.7 on a Likert Scale is about as meaningful as a 'fair and a half'. Non-parametric data make fewer assumptions about the normality of the distributions of the populations being analysed and about the sample sizes of the groups being compared (Green & Salkind, 2008). Central tendencies, using non-parametric statistics, are expressed in terms of median (middle value) or mode (the most frequently occurring response to points in the scale).

Having considered the arguments it was decided that the use of non-parametric procedures was more appropriate for the purposes of this study.

3.3.4.4 Statistical analysis tests for the present study.

3.3.4.4.1 The Kruskal-Wallis test.

The present study aimed to examine the relationship between certain background characteristics of the learner evaluators and their scoring on the Likert scale-based questionnaire items. These background factors (detailed in Chapter 5.9.2) are:

- gender
- pupils' year in school
- pupils' school type (Gaeltacht/Gaelscoil/English-medium school)
- the frequency with which pupils play computer games generally
- pupils' perception of their own ability level at Irish
- pupils' rating of their 'Top 3' favourite subjects
- pupils' preconceived attitude towards TTS synthetic voices in general

This information was elicited before learners explored the platforms and is referred to in this study as 'pre-game questionnaire items'.

Learners' responses to the Likert scale-based questionnaire items were elicited after they had experienced the platforms and they are referred to in this study as 'post-game questionnaire items' (detailed responses are given in Chapter 5.9.3).

The statistical analysis technique used for determining the relationship between these background factors (pre-game questionnaire items) and learners' responses to post-game questionnaire items in the present study is the Kruskal-Wallis test. This technique is used in non-parametric statistics to test for differences in ranks among three or more groups (Cohen, 2001). Kruskal-Wallis is the non-parametric alternative to a one-way between-

groups analysis of variance (Pallant, 2010). Groups here are based on the independent variable, say, school type or the point at which pupils placed themselves on the Likert scale of opinions, e.g. attitude to synthetic speech generally. The Kruskal-Wallis analysis looks for significant differences between these groups in their responses to the various post-game items which are being regarded here as dependent variables.

The Kruskal-Wallis test is a generalisation of the Mann-Whitney *U* test. The former can be used for three or more groups whereas the latter is confined to two groups. The test is based on an analysis of the rankings rather than on means and looks for significant differences between the rankings assigned by the respondents to the various dependent variables. The tests are less highly influenced by outliers, or extreme scores, since any score is but a ranking and any extreme score is regarded as being just one rank below the previous highest or lowest score (Cohen, 2001). The test allows one to examine, for example, if there is a significant difference between the rankings on a Likert scale which respondents gave to the quality of the graphics based on the level of their like/dislike of synthetic speech generally.

When testing for statistically significant between-group differences, the Kruskal-Wallis test merely highlights that there is a difference between the groups. It does not, however, pinpoint where the significance lies. There may, for example, be a significant difference between Group A and Group C but the difference between Group A and Group B or Group B and Group C may not be statistically significant. In order to determine where exactly the difference lies, a further post-hoc test must be carried out. The Mann-Whitney *U* test is commonly used for this post hoc analysis. The Kruskal-Wallis test, however, is but an extension of the Mann-Whitney *U* test and produces identical results for two independent samples (Green & Salkind, 2008, p. 390).

3.3.4.4.2 Spearman's Rank Correlation Coefficient.

Spearman's Rank Correlation Coefficient, also known as Spearman's rho, is a technique which calculates correlations and is used for non-parametric data. Its parametric equivalent is the Pearson Product-Moment Correlation. The Spearmans' rho is used with ordinal or ranked data (Pallant, 2010). The reason for using it in the present study is to establish the correlation between responses to the various post-game questionnaire items in order to examine the extent to which one can claim that the respondents' experience of the CALL platforms is a holistic one, be that a favourable holistic experience or an unfavourable one.

Like the Pearson correlation coefficients, the Spearman's rho coefficient takes on a value varying from -1 to +1. The strength of the relationship between the two measurements is determined by the correlation coefficient. A coefficient of 1 shows that for every change in A there is an exactly equivalent change in B. A coefficient of -1 shows that for each change in A, an equivalent change, in the opposite direction, is experienced by B. As A increases, B decreases by an exactly similar amount. Cohen (1988) gave the following guidelines for interpreting correlation coefficients:

Small correlation: r = .10 to .29

Medium correlation: r = .30 to .49

Strong correlation: r = .50 to 1.00

As with the Kruskal-Wallis test, the Spearman's rho coefficients are also based on rankings rather than on actual scores. Results from the Spearman's rho test are generally presented in grid format so that one can readily perceive the strength of the relationship between the dependent and independent variables. In this case, one can readily see the relationship between responses to all post-game questionnaire items and judge the extent to which the platforms were regarded by the respondents as 'holistic' experiences. A high

number of statistically significant relationships between the items in the post-game questionnaire would suggest that there is a significant relationship between respondents' reactions to the various items. If respondents have a somewhat similar attitude, be that favourable or unfavourable, to most of the post-game questionnaire items, then one can claim that the overall experience was a holistic one, i.e. respondents have a similar degree of positivity or negativity towards most of the items. If, for example, respondents were relatively unfavourable towards one item, and there was a significant positive correlation between that item and all the other items, then one can claim that they were likely to be relatively unfavourably disposed towards all the other items. The opposite also applies. In carrying out the Spearman's rho test, one is testing for statistically significant correlations but this does not show whether or not the respondents had a positive or a negative attitude towards the entire game. This test is carried out on the learner data in Chapter 5.9.3.3.6.

3.3.5 Summary of evaluation criteria.

This section has concerned itself with a review of the methods appropriate for the evaluation of CALL programs in general. Section 3.3.1 began with a more general review of the literature on the evaluation of CALL materials, followed by specific criteria for evaluating the CALL platforms developed as part of the present project.

A distinction between objective and subjective data is made and the opinion of previous researchers that subjective data on their own is insufficient, is noted. Subjective data refers mainly to opinions expressed by some stakeholders in relation to the CALL program. Objective data, on the other hand, are obtained from some observable behaviour and are generally regarded as being more reliable. They are frequently acquired by quasi-experimental methods whereby a population is divided into two groups, one of which receives the treatment while the other does not. At the end of the treatment period any differences between the two groups in relation to the predefined criteria are ascribed to the treatment. It is argued here that this method is inadequate for the evaluation of CALL platforms. One cannot directly attribute some language structure which a learner is found

to have solely to the CALL platform which they have experienced. Language acquisition is not an absolute in that it has an absolute presence or an absolute absence. We acquire language to various degrees and may understand or use words or phrases correctly in some contexts but incorrectly in others. It was concluded that it may not be useful to seek general criteria for the evaluation of CALL but rather, that each CALL platform must be evaluated according to its own terms, i.e. its aims and objectives, the group to which it is being directed and the teachers or administrators of the program. It is from this context that the criteria for the evaluation of the present platforms are being drawn up.

Section 3.3.2 reviewed the criteria by which TTS synthetic speech may be evaluated beginning also with a more general review first before developing the criteria for evaluating the Irish TTS synthetic voices. While a good deal of work on general evaluations on TTS synthetic speech has been done, no previous evaluations have been carried out in the context of game-like CALL activities. It is recognised that evaluations of synthetic speech should be done in the context where the speech is expected to be used. It is appropriate, then, that the Irish synthetic speech of *ABAIR* be judged in the context of three CALL platforms. This evaluation is considered to be quite different from a general evaluation of the output of the *ABAIR* system.

The concept of 'ease of intelligibility' was proposed as a criterion for evaluating synthetic speech. A detailed consideration of a formula by which assessment of 'intelligibility' and 'ease of intelligibility' of sentences may be combined so as to produce an overall *Clarity Index* is given in Chapter 4.3.2.2.2 under the heading "Developing a Clarity Index for synthetic speech".

3.4 Statement of Hypotheses

A formal statement of research questions is made in this section in the form of specific hypotheses. These hypotheses are an amalgam of the specific research questions which this study sets out to investigate and deductions drawn from the reviews of literature presented in Chapters 2 and 3 which form the basis for the research questions. Some of these hypotheses lend themselves to scientific statistical testing of probability while others are such that support for them is shown by way of trends in the data, such as percentages of respondents who accept or reject a particular proposition.

The main research aim, as set out in Chapter 1.1 of this thesis, is to examine the contribution which Irish synthetic speech can make to the development of future CALL platforms. This aim is broken down into a number of specific objectives, each of which may be examined in turn. Each of these objectives is stated here in terms of a hypothesis.

Hypothesis 1:

Digichaint, as an interactive game-like platform, will conform to the criteria for effective CALL platforms (as set out in Sections 3.3 above, see Table 3.2):

1.1 - game is task-based with a focus on meaning
1.2 - game has an appropriate language difficulty level
1.3 - game is easy to manoeuvre
1.4 - game is enjoyable
1.5 - game has a clear plot
1.6 - game has a logical structure
1.7 - game is engaging for the learners
1.8 - game has graphics that are considered attractive/engaging
1.9 - game has legitimate narrative, i.e. credibility for the player/learner

Hypotheses 2, 3 and 4:

In terms of synthetic speech, one needs to establish the respondents' views on its intelligibility, quality and attractiveness. Hypotheses 2, 3 and 4 are central to the main research aims and objectives of the present study, as set out in Chapter 1.1. They are key to the research objectives in that they involve the basic tests of adequacy and 'fitness-for-purpose' of synthetic voices in the context of game-like CALL platforms.

Hypothesis 2:

The synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability.

Hypothesis 3:

The synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness.

Hypothesis 4:

A majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented.

Hypothesis 5:

There is a statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms (such as plot, ease of manoeuvring characters, legitimacy of narrative and standard of Irish language used) and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform.

Hypothesis 6:

There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general and their attitude towards the Irish synthetic voices being tested.

3.5 Summary

The current chapter begins by detailing the development of three CALL platforms which form the basis for the current research project. It then reviews the literature on the criteria for evaluating CALL platforms and develops the specific criteria by which the present platforms will be evaluated. It also reviews the literature on the evaluation of TTS synthetic speech and, similarly, identifies the criteria by which the synthetic speech, which is integrated into the present platforms, may be evaluated. On the basis of this review of literature and of the review of literature, specific hypotheses are developed and these hypotheses encapsulate the criteria which are deemed to be the ideal for effective CALL platforms with integrated synthetic speech.

The platforms, described in Section 3.2, have been named *'Digichaint', 'Fáilte go TCD'* and *'Taidhgín'* and each has its own unique features. *Digichaint* carries the features which are most closely associated with computer gaming in that the player has a particular problem to solve and must make use of the resources available to him/her from within the world of the game in order to do so. *Fáilte go TCD* has highly stylised graphics but is low in interactivity insofar as the learner has little control over the direction of the 'game'. *Taidhgín* is highly interactive with relatively simple graphics. It is a conversation partner or 'chatbot' which can process novel utterances inputted by the learner and respond in a linguistically appropriate manner. The synthetic speech which is being developed as part of the *ABAIR* project was integrated into each of the platforms and an essential part of the present thesis is the evaluation of the speech in the context of the three CALL platforms.

A review of the literature on the evaluation of CALL materials (see Section 3.3.1) identifies the many facets which are considered to be relevant. The most obvious of these is the potential CALL platforms offer to learners to enhance their learning experience. Evaluating this potential can be problematic since it is difficult to credit a particular CALL platform with being the sole provider of linguistic input to the learner. Even the concept of acquiring a particular language form is problematic since there are very many levels of acquisition ranging from a vague acquaintance with the form to a full knowledge of the entire range of environments in which it may be used. It is argued that the test – treatment - retest model of research is inappropriate for CALL evaluations since changes observed from test to retest cannot be readily assigned to the 'treatment' or exposure to the CALL program. An evaluation of CALL materials must concentrate on the quality of the engagement of the learner with the materials. This can be done by observing the learners interacting with the materials and/or seeking the learners' opinions on their experiences. This chapter contains a list of nine criteria, which have been deduced from the literature, by which one may evaluate CALL platforms and these criteria form the basis for the hypotheses formalised in Section 3.4.

Section 3.3.2 of the chapter concentrated on the evaluation of the TTS synthesis as produced by the *ABAIR* project. There has been no previous evaluation of this synthetic speech as it is still at a developmental stage and the present study does not claim to represent a comprehensive evaluation of the speech. Rather, it sets out to evaluate the adequacy of the synthetic speech for interactive CALL platforms. To my knowledge, no previous evaluation of TTS synthesis for any language in such a context has been carried out.

Synthetic speech in an interactive CALL platform is important because it allows for novel interactions between learner and computer. With the use of Artificial Intelligence

(AI), chatbots can hold relatively realistic conversations with the learner. The use of pre-recorded speech is not possible since the utterances are unpredictable.

In the context of the evaluation of the present TTS synthesis, the functionality, or fitness-for-purpose, of the speech was considered paramount. The three evaluation criteria identified as being most relevant in the present context were intelligibility, quality and attractiveness. It was decided that there should be two main groups of evaluators:

(i) those with a native or near-native-like Irish language ability. A group of Irish language teachers was chosen as representing this group

(ii) school pupils who were in the same age bracket as the target group for whom the platforms were being developed. Pupils were taken from three different school types:

   a. Gaeltacht schools were Irish is a community language
   b. Gaelscoileanna, which are located in English speaking areas of the country but who teach all subjects through the medium of Irish
   c. English-medium schools where Irish is but a subject in the curriculum

There needed to be a heavy weighting in favour of schools with an Irish speaking background since one was attempting to ensure that any intelligibility deficiencies were due to deficiencies found in the synthetic speech itself rather than weaknesses of the language comprehension ability of the listeners. Since it was envisaged that the end-users of these platforms would ultimately be learners of Irish typically in English-medium schools, a representative group was also included so as to get their evaluations on all the key aspects of the platforms including the synthetic speech (TTS) and the non-synthetic speech (non-TTS) features.

Section 3.3.4 reviews the statistical methods used for the analysis of the data obtained from the evaluators. The use of Likert scales for the generation of the data is discussed as

are the concepts of parametric and non-parametric data. It was decided that the data generated from the Likert questionnaires should be treated as being non-parametric. This determines the statistical techniques which can be used for analysis. The Kruskal-Wallis technique was chosen as being the most appropriate to test the relationship between the background characteristics of the respondents, such as gender, level of understanding of Irish, etc. and their evaluations of the various aspects of the platforms. The test searches for significant between-group differences in their responses to the Likert scale-based questionnaire items.

A brief description of Spearman's Rank Correlation Coefficient is also given. This technique is used to establish the correlation between the responses given to the various items on the Likert scale-based questionnaire in order to test the proposition that the use of any of the platforms is a holistic experience and that one's views on any aspect of it will be influenced by one's general experience of the whole. In such a scenario it may be claimed that one's assessment may be somewhat dependent on the context in which it is used.

The discussion to this point (in both Chapter 2 and Sections 3.2. and 3.3) leads to a formal statement of hypotheses in Section 3.4, which draws together those aspects of the literature reviews most pertinent to the present study. Chapters 4 and 5 of this thesis describe the procedures used for the generation and analysis of the data which test the hypotheses stated in Section 3.4 of the present chapter. Chapter 4 concentrates on the data pertaining to teacher respondents and Chapter 5 deals with the data from learner respondents. It is postulated that one's language comprehension ability may be a significant factor in determining one's judgment of synthetic speech. In this context it was important to get a wide Irish language ability range amongst evaluators in order to test this proposition.

Chapter 4: Findings for Teachers

4.1 Overview

The main focus of the overall study is to develop and evaluate CALL platforms and to assess the *ABAIR* TTS synthetic speech in these game-like CALL contexts. Specifically, this chapter sets out to address Hypotheses 2, 3 and 4, as stated in Chapter 3.4. Hypothesis 2 refers to the intelligibility of the synthetic speech for those with a high Irish language comprehension ability. Hypothesis 3 refers to the quality and attractiveness of the synthetic voices and Hypothesis 4 refers to their acceptability in the context of CALL platforms. Two main groups used as evaluators – teachers and learners. The present chapter discusses the methodologies used and the results acquired from the teacher section of the evaluation.

For the teacher evaluators of the TTS synthetic voices it was decided to use an orthographic transcription task in order to assess the intelligibility of the voices. It is assumed that teachers' comprehension of Irish is at a very high level and any deficiencies in their transcriptions could reasonably be ascribed to deficiencies in the synthetic voices rather than deficiencies in the linguistic ability of the evaluators. This assumption could not be made in respect of the learners. One could not make judgments on deficiencies in the transcriptions of such learners with any certainty, as deficiencies could be ascribed either to the synthetic voices or to the learners' restricted linguistic capacity. It was under these circumstances that it was decided that teachers should be used in order to elicit what can be described as 'objective' data for this study, through an orthographic transcription task. The opinions of teachers were also sought (described as 'subjective' data) and this chapter contains a model for integrating the results of the subjective and objective data. It was decided to elicit the opinions of learners separately (see Chapter 5) and the data acquired from that section of the study can be described as being 'subjective' (see Chapter 1.4 'Definition of Terms').

In using both teachers and learners as evaluators in the overall study one is attempting to assess, in both objective and subjective terms, the overall value of the three CALL platforms, with integrated synthetic speech, which were developed as part of the present study. Focus in this chapter is kept mainly on an evaluation of the synthetic speech. Data for the evaluation are taken from the *Digichaint* platform but the teachers were not asked to play the game as part of their evaluation. It must be noted that although the material for the evaluation was taken from the *Digichaint* platform the voices used were all in their original synthetic forms, and versions of the voices with pitch and speed manipulations, as described in Chapter 3.2, were not used. Although teachers were shown *Fáilte go TCD* and *Taidhgín* in an informal context, no formal evaluation was carried out and teachers' opinions on the TTS and CALL aspects of those two platforms have not been included in the present thesis. Chapter 5 focuses on learner evaluations, where one is simultaneously evaluating the acceptability of the platforms as a whole, which includes its linguistic (i.e. synthetic voices) and non-linguistic aspects (i.e. graphics, etc.).

The present chapter deals specifically with teachers' reactions to the TTS synthetic voices in the context of the *Digichaint* game (see Chapter 3.2.1). The chapter includes an orthographic transcription task devised for teachers in order to assess the 'intelligibility' of the voices and it also measures the opinions of the teachers in relation to the 'ease' with which they could transcribe the sentences. It contains a discussion on the concepts of 'intelligibility' and 'ease of intelligibility'. The performance of the teachers on the orthographic transcription (i.e. intelligibility) task will be referred to in this chapter as a 'performance' score. The ratings given by teachers on the 'ease of intelligibility' task are referred to as 'opinion' scores. The 'performance' score may be seen as an objective score since it involves an objective procedure for marking the performance of the teachers in transcribing the twenty sentences. The 'opinion' score is considered a subjective measure as it involves teachers marking their opinions on the 'ease' with which they performed in the transcription task. A formula is devised which combines the 'performance' and 'opinion' scores of the evaluators for each of the twenty sentences

included in the task. A 'clarity' index is proposed which combines the 'performance' and 'opinion' scores. Clarity here refers to a combination of 'intelligibility' and 'ease of intelligibility'. The relationship of the various components of the chapter is outlined in Table 4.1 below.

Table 4.1

The constituents of the proposed 'Clarity Index' - combining 'intelligibility' (performance scores) and 'ease' of intelligibility (opinion scores).

Aim	**Task devised to Elicit Data** (described in Section 4.3.2)	**Nature of Data**	**Calculation** (Sentence analysis: see Section 4.3.2.2.1)	**Formula** (*Clarity Index*: see Section 4.3.2.2.2)
Intelligibility	Orthographic Transcription Task	Objective Data	*Performance* score	**Clarity Index**
'Ease' of intelligibility	Opinion Rating (1-5: Likert Scale)	Subjective Data	*Opinion* score	

The Design and Results sections of the present chapter (see Sections 4.3 and 4.9, respectively) are divided into four main sections which detail:

(a) pre-test questionnaires, which are designed to elicit background information on each of the participants (see Section 4.3.2.1 and Section 4.9.2)

(b) the test (see Sections 4.3.2.2 and 4.9.3), which includes:

   a. a sentence analysis task, which comprises both an orthographic transcription task (which generates a '*performance* score') and an '*opinion*' rating of the 'ease' with which that transcription was done ('*opinion* score')

b. a *Clarity Index* (a formula devised to combine teachers' 'performance' scores with their 'opinion' scores (i.e. 'intelligibility' and 'ease of intelligibility' scores)

(c) post-test questionnaires, which elicit teachers' opinions on the usefulness, quality and attractiveness of the *ABAIR* synthetic voices (see Section 4.3.2.3 and Section 4.9.4)

4.2 Participants

It was decided to use teachers as evaluators of the synthetic speech because of their high competence in the Irish language. It could be assumed that a lack of understanding of the TTS speech would be due to deficiencies in the quality of the synthetic speech rather than deficiencies in the evaluators' language comprehension ability. It was decided that the data for assessing TTS intelligibility would be generated by asking the teacher participants to orthographically transcribe 20 sentences, each of which was to be found in the *Digichaint* platform.

When the sentences for transcription were decided upon (see Section 4.3.2), an initial pilot study amongst five second level teachers was conducted. These showed largely positive results and did not justify any adjustments to the original selection of sentences. Data from these five teachers were not included in the final analysis of results.

Invitations to take part in the present study were extended to teachers who attended an Irish language teaching conference in NUIM, Maynooth (An Tionól Teagaisc, 2011). A large group of teachers showed enthusiasm for the study on the day. A total of 31 respondents with no reported hearing impairments participated in the final evaluation. The group was predominantely female (N=25). Participation was voluntary and not organised, in the sense of participants assembling in one place at one time. The evaluations were conducted online when the participants found the time to partake. A

detailed description of the composition of the teacher group that participated in the project is given in Table 4.4, Section 4.9.2 below.

4.3 Teachers: Research Design and Strategy

4.3.1 Research Design.

This section of the research, which focuses solely on teacher evaluations, involves a 3-stage multi-method design (Creswell, 2009) comprised of three instruments: a pre-test questionnaire, a test and a post-test questionnaire.

Stage 1	Stage 2	Stage 3
Pre-test Questionnaire → Elicits background information from the participants	**Test** → Orthographic transcription task	**Post-test Questionnaire** Seeks teachers' general reaction to the TTS in the context of a CALL game-like platform

Figure 4.1. Research Design for Teacher Evaluation

The design for this part of the study involves:

(a) Stage 1: collecting background information on the participants. This was done by way of a pre-test questionnaire

(b) Stage 2: administering the test items, which were finally decided upon subsequent to the administration of an initial pilot study

(c) Stage 3: a post-test questionnaire which got teachers' general reactions to the TTS voices in the context of a CALL game-like platform

The pre-test questionnaire involved eliciting certain background information about the evaluators, which may have a bearing on the response patterns of different subgroups (see Section 4.3.2.1).

The 'test' in this part of the study involved the evaluators giving an orthographic transcription of twenty sentences taken from the *Digichaint* platform. This transcription task was necessary in order to investigate the intelligibility of the synthetic speech as well as participants' opinions on the ease with which they could carry out the transcriptions. The rationale for using the particular sentences chosen for transcription is outlined in Section 4.3.2.2.1, as is the marking scheme through which teachers were assigned 'performance' scores. Section 4.3.2.2.2 describes the development of a formula which is devised in order to produce a *Clarity Index* that combines 'performance' scores and 'opinion' scores.

A short post-test questionnaire which asked for teachers' more general opinions on the usefulness, attractiveness and quality of the *ABAIR* synthetic speech (see Section 4.3.2.3) was also included.

4.3.2 Research Strategy.

4.3.2.1 Pre-test questionnaire.

Background information was sought from each participant in relation to his/her general age group, gender, dialectal preference, and preconceived views of synthetic voices in general. It was anticipated that these factors may lead to between-participant variations. Different groups may have difficulty in comprehending dialects apart from their own (see Ó Flatharta (1989), Chapter 3.3.3) and other factors such as age may have an effect on listeners' willingness to accept the synthetic voice. The pre-test questionnaire used in the study is included in full in Appendix B1.

4.3.2.2 Test.

The test material was all collated into one electronic form, consisting of four webpages, and was presented to evaluators online. Participants were presented with a 3-minute video consisting of an excerpt from the *Digichaint* game. This was intended to acclimatise participants to the synthetic speech and to give them a context to better understand the purpose of the research. The overall aim of the test was to evaluate teachers' performance and opinions on an '*intelligibility*' and '*ease of intelligibility*' task. This section has been divided into two subsections – 4.3.2.2.1, which details the design of the orthographic transcription task (i.e. the content and the criteria for marking teachers' transcriptions) and 4.3.2.2.2, which describes the development of a formula to create a *Clarity Index*, combining 'performance' scores and 'opinion' scores.

4.3.2.2.1 Sentence analysis task: evaluating 'intelligibility' and 'ease'.

In order to elicit teachers' evaluations of the TTS synthetic voices, a technique used by Hashimoto, Yamagishi, Byrne, King and Tokuda (2011) was employed, i.e. the teachers were asked to listen to individual sentences and to orthographically transcribe what they heard. These transcriptions were then assessed for accuracy. Teachers were asked to simultaneously give a rating on the 'ease' with which they could transcribe these sentences on a scale of 1 to 5, reflecting their opinion of the clarity of the sentence.

Twenty utterances were taken from the overall corpus of speech used in the *Digichaint* game (ten of Connaught dialect and ten of Ulster dialect). The sentences appeared in random order each time the test was given, so that their position in the list would not influence their intelligibility. If a particular sentence were to always appear last, for example, it could get a poorer rating because of participants' fatigue or inattention.

The twenty utterances were divided into five categories, each containing four utterances – two utterances in the Ulster dialect (female voice) and two utterances in the Connaught dialect (male voice), and these are presented in Appendix B3. The utterances selected for

the teacher orthographic transcription test were not random, but were chosen so as to include a representative selection of sound classes in the Irish sound system (long vowels; short vowels; fricatives and stops) (Ó Siadhail, 1980) as well as samples of elision in fixed common phrases. Fricatives and stops in particular are included as they can give rise to some production problems during the development phase of synthetic voices. The five categories of utterances tested are outlined in Table 4.2 below.

Table 4.2

The *five categories used when selecting sentences for inclusion in the teacher orthographic transcription task*

1.	utterances featuring elision (**e.g.** "Tá, an bhfuil a fhios agat cá bhfuil Tom Bán" – pronounced: **with elision**: t̪ˠ a: vˠ i lʲ i sˠ a t̪ˠ k a: lʲ t̪ˠ o mˠ bˠ a: n̪ˠ **without elision**: t̪ˠ a: ə n̪ˠ vˠ i lʲ ə i sˠ ə g ə t̪ˠ k a: vˠ i lʲ t̪ˠ o mˠ bˠ a: n̪ˠ **Translation**: *Yes, do you know where Tom Bán is?*)
2.	utterances featuring four or more long vowels - with a long vowel being at the focal point of the utterance
3.	utterances featuring four or more short vowels - with a short vowel being at the focal point of the utterance
4.	utterances featuring three or more fricatives
5.	utterances featuring three or more stops (plosives)

** In the circumstances of the present project, I had access to the back end of the synthesiser and was able to include a lexicon containing commonly-used phrases found in the game which are normally pronounced as one word (following elision) rather than having each component word pronounced individually (e.g. cá bhfuil... /*k a: vˠ i lʲ/ -> *pronounced /k a: lʲ/ (*Translation: *Where is...)).*

Criteria for marking responses to 'intelligibility' test.

As teachers transcribed each sentence, they were also asked to rate on a 5-point Likert scale the 'ease' with which they transcribed that particular sentence (with "5" denoting complete ease of transcription and "1" denoting very difficult to transcribe). It was intended that this would add robustness to the measure of the intelligibility of the voices. Even if a particular sentence was transcribed completely correctly, a low rating on the 'ease of intelligibility' scale would indicate that the voices were of a lower quality than the mere transcription mark alone would indicate. The ratings given for 'ease of intelligibility' are reported as subjective data. The twenty sentences used in the test are listed in Appendix B3.

The results from the orthographic transcription task are reported here as objective data. Marks between 1 and 5 were assigned to each of the transcribed sentences in the following manner:

Table 4.3

Ratings awarded to teachers based on the orthographic transcription task

5.	The transcription is fully correct
4.	The transcription is largely correct but contains minor inaccuracies that do not affect the meaning of the sentence – e.g. "ok, cé hé féin" in place of "ó, cé hé féin"
3.	The category being examined (see Table 4.2) is transcribed correctly but other words in the sentence are transcribed incorrectly which may lead to a change in meaning – e.g. in Category 5 (Stops) / Ulster dialect: "tá turas **i gcroí** duine inniu" (t̪ʸ a: t̪ʸ u ɾʸ ə sʸ **i g ɾʸ i:** d̪ʸ i nʲ ə ə n̪ʲ u) (*there's a trip in someone's heart today*) in place of "tá turas **eagraithe** dúinn inniu" (t̪ʸ a: t̪ʸ u ɾʸ ə sʸ **i g ɾʸ i:** d̪ʸ u: n̪ʲ ə n̪ʲ u:) (*there's a trip organised for us today*) The incorrect representation is shown in bold type
2.	The category being examined is transcribed incorrectly - thus an error in a fricative in the 'fricative sentence group' resulted in a rating of '2' – e.g. "Ar éigean … an doras anois" in place of "Ar éigean go n-aithneoinn é dá siúlfadh sé isteach an doras anois" (*Hardly would I recognise him if he were to walk in the door now*)
1.	The answer is missing or not discernable

It is taken here that these marks form a scale, rather than representing ordinal ranking in that a score of '5' represents a transcription which is superior in quality to that of a '4', which in turn is superior in quality to that of a '3', etc. Since '5' represents the ideal / perfect transcription, it denotes a high quality in the synthetic voice.

4.3.2.2.2 Developing a Clarity Index for synthetic speech.

Performance alone is not a clear indicator of the intelligibility of utterances. Utterances in particular contexts may be intelligible but require considerable effort from the listener. In the present study, it was considered desirable not only to ask respondents to transcribe sentences (as a measure of 'intelligibility') but then to enquire into the 'ease' with which they felt they could do this ('ease of intelligibility'). This section describes a formula which was newly devised as part of the present study in order to combine both these measures. A mean score, representing both 'performance' and 'opinion', is calculated for each sentence. An overall mean of the respondents' 'performance' scores over twenty sentences is calculated and similarly an overall mean of their 'opinion' scores is calculated. These means are inserted into the formula below.

P_1 represents the mean performance score over 20 sentences for respondent 1. O_1 represents the mean opinion score over the 20 sentences for respondent 1. "5" represents an ideal score where the respondent can transcribe the sentence with complete ease. The gap between an actual score for a particular sentence and the ideal score represents the difficulty level of that item. The actual score is a combination of the performance score and the opinion score. The score for an item having a performance score of, say 4, and an opinion score of 2 may be seen as $(5-4)+(5-2)$. The mean for the deviation of performance and opinion scores from the ideal 5 is therefore $\frac{1+3}{2}=2$.

The extension of this operation to cover all respondents may be represented as follows:

$$5 - \frac{\sum_{n=1}^{N}(5-P_n)+(5-O_n)}{2n}$$

N = the total number of respondents

P_1 = the mean performance score of respondent 1

O_1 = the mean opinion score of respondent 1

Figure 4.2. Clarity Index formula

The denominator used to calculate the mean is 2n since each respondent has two scores – a performance score and an opinion score. Results of applying this formula are presented in Section 4.9.3.2 below.

4.3.2.3 Post-test questionnaire.

Having completed the test, participants were asked for their opinion on the suitability of the synthetic speech for use in games such as *Digichaint* ("*Do you think synthesised speech at this level would be suitable for a game like Digichaint?*"). They were also asked for their opinion on the attractiveness of the voices for senior cycle pupils ("*Do you think the synthesised speech in the sample you heard would be attractive for senior cycle pupils in second level schools?*") and for their general opinion of the standard of the synthetic speech ("*What is your opinion of the quality of the speech produced by abair.ie? (Please rate on a scale of 1 – 5)*"). Respondents were given five statements in relation to the questions pertaining to suitability and attractiveness and asked to pick the one which best reflected their opinions. They were also given an opportunity to give general feedback, be that positive or negative, in an open dialogue box. The post-test questionnaire, as it was presented to teachers, is included in Appendix B4.

4.4 Instruments

As described in Section 4.3, responses from a pre-test questionnaire, a test and a short post-test questionnaire were elicited from teachers. This was all carried out by way of online evaluations for which participants volunteered to take part at a time and in a place of their own choosing.

The pre-test questionnaire (see Appendix B1) sought to establish the identity of each of the evaluators and some details of their backgrounds. These identity tags were necessary in order to organise the data and respondents were free to create their own username, which may or may not be their actual name. The information being sought referred to gender; age-group; place of birth; dialect preference; and teachers' pre-conceived ideas about synthetic voices generally. There has been little previous research on gender variations in responses to synthetic speech (Crabtree, Mirenda, & Beukelman, 1990; Lee, 2008), and few studies which investigate how different age-groups may respond to synthetic speech (Drager & Reichle, 2001). Intuitively, it would seem that there may be variations associated with age-groups since the younger age-groups would possibly have a greater familiarity with digital media, which may include synthetic voices. A question on place of birth was included in order to identify native and non-native speakers of Irish. It is possible that native speakers of Irish from one Gaeltacht area may have some difficulty in understanding certain sounds/phrases from another Gaeltacht area. A question on teachers' preferred dialect was included, mainly in order to establish if those who were born outside Gaeltacht areas had a particular dialect preference and whether this may influence their attitude towards the dialects included in the present test. A question on teachers' pre-conceived ideas about synthetic speech was included in order to test the proposition that one's predisposition towards synthetic speech influenced one's evaluation of it (see Cryer and Home (2009) and discussion in Chapter 3.3.2.1).

Evaluators were asked to orthographically transcribe twenty sentences which had been included in the *Digichaint* platform. They were first given a demonstration of the

Digichaint platform, by way of a 3-minute video of the game, in order to familiarise them with the synthetic voices and clarify for them the context in which the synthetic speech would be used. The sentences that were included in the transcription task were not contained in the introductory video. They were all in their original synthetic forms – none of the pitch and speed manipulations that had been added to these voices to make the characters in *Digichaint* sound different from one another were included in this test set. The sentences were presented in random order on the test webpage which gave them a 'stand-alone' status – i.e. one could not glean a sentence from the broader pragmatic context. The same webpage contained the 'ease of intelligibility' rating scale, which was presented alongside each of the twenty sentences.

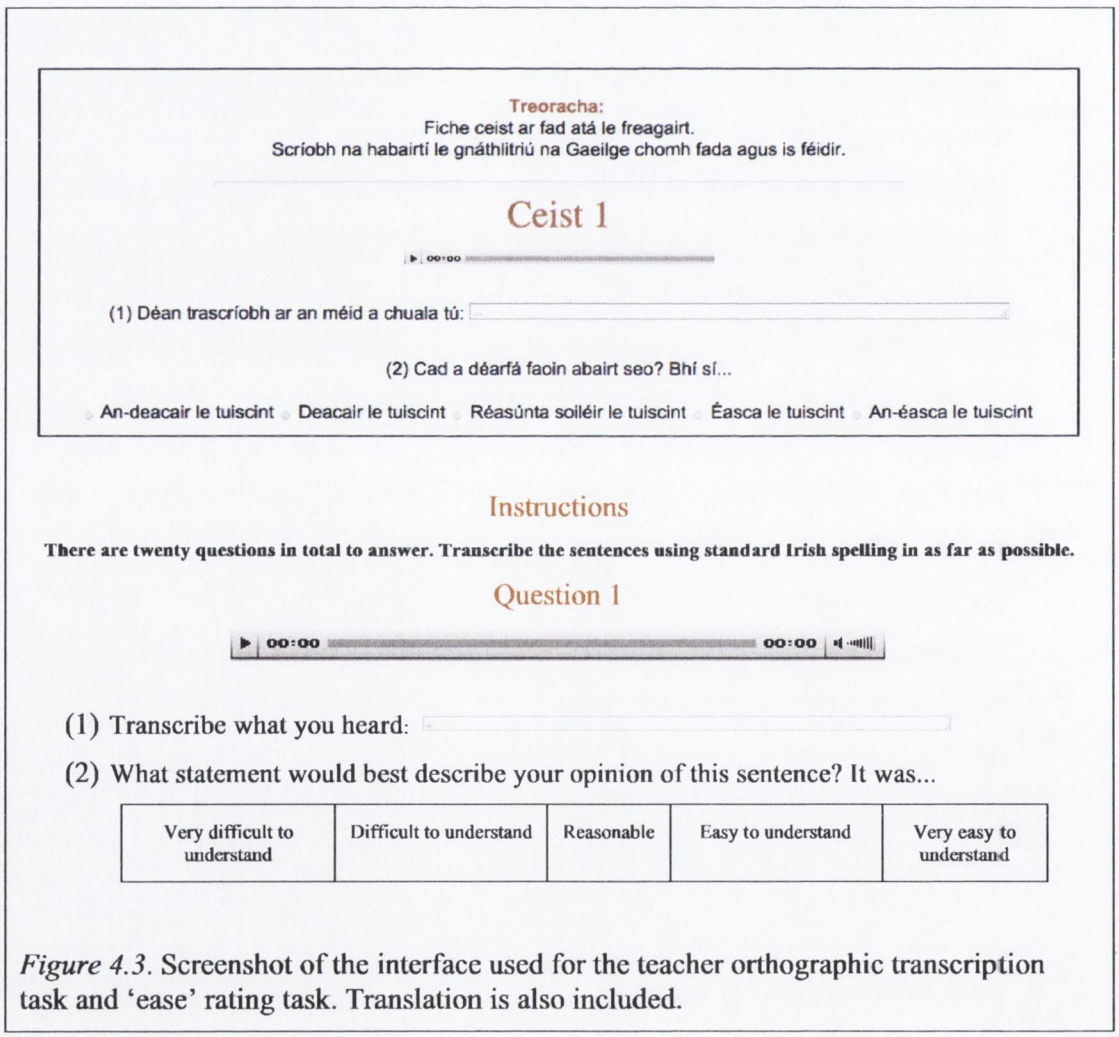


Figure 4.3. Screenshot of the interface used for the teacher orthographic transcription task and 'ease' rating task. Translation is also included.

A post-test questionnaire, which asked three questions and gave teachers the opportunity to give feedback, be that positive or negative, through an open-ended dialogue box, was included. The first question sought to establish the suitability of the synthetic speech for a game like *Digichaint*. These opinions were important in the study since they would inform on the suitability of the test for the subsequent learner evaluation. The second question sought teachers' opinions on the attractiveness of the voices for second level learners of Irish (see discussion in Chapter 3.3.2.2). The third question was more general and it sought to establish the views of the teachers on the overall standard of the *ABAIR* synthetic speech. This was included so that one could glean an overall positive or negative attitude towards synthetic speech among the teacher population. An open dialogue box was included to allow for any unanticipated features of the synthetic speech or the game which may have come to the attention of teachers. It offered a forum in which the teachers could communicate their general reaction to the platform and the synthetic voices since I was not present while the teachers took part in the evaluation and therefore could not discuss issues with them at the time of their participation.

4.5 Procedure

The teachers' evaluation was conducted online where individuals volunteered to participate in the evaluation. They did this at a time and place of their own choosing and did not have any contact with each other. I advised them to use Google Chrome as their browser and to use headphones to listen to the synthetic voices at a suitable volume. I cannot, however, be certain that optimal conditions were available to all those who participated. For each respondent there was a pre-test questionnaire, followed by the test, and a post-test questionnaire.

As outlined in Section 4.3.2.2 above, respondents were asked to listen to twenty sentences altogether, (10 in the Connaught dialect and 10 in the Ulster dialect), and to transcribe them as they heard them. They were simultaneously asked to rate the 'ease' with which they could transcribe the sentences on a Likert scale which appeared with the

soundfile of each of the twenty utterances. Teachers were free to listen to the sound files as many times as they wished. These sentences are included in Appendix B3.

4.6 Analysis

Data from the pre-test questionnaire are presented in Table 4.4 below in a purely descriptive manner. The table lays out in percentage terms the backgrounds of the evaluators who participated in the study.

The scores given by the evaluators to each of the twenty sentences that make the 'test' are set out in Appendix D1 in terms of mean, median and mode. These scores are presented under the five sentence-categories described in Section 4.3.2.2.1 above – i.e. elision; long vowels; short vowels; fricatives; and stops. These categories were selected because they are features of the language and because some of them (fricatives, stops and elision in particular) could be considered potentially problematic. It was deemed necessary to include mean, median and mode scores in the table in Appendix D1 in order to clarify measures of central tendencies and in order to avoid having the central tendency over-influenced by a few extreme scores. One needs to be particularly careful of this problem when the number of respondents is relatively low (in this case N=31). For the purposes of demonstration here one may refer to 'Sentence analysis findings' (see Section 4.9.3.1). The mean objective evaluation score for Item 14, for example, is 3.2 but its median and modal scores are 4. This shows that the mean score was over-influenced by a few low scores but that most respondents found no particular difficulty with the item. By contrast, Item 17 also had a mean score of 3.2 but its median and modal scores were 3, which indicates that a far greater number of respondents had difficulty with the item than was the case for Item 14. A more detailed discussion of the scores on the various test items follows in Section 4.9.3.1.

The formula by which the *Clarity Index* is calculated is set out in Section 4.3.2.2.2 above. Data from the calculation are set out in Appendix D2.

Responses from the three questions posed in the post-test questionnaire are set out in descriptive terms in Table 4.7 below.

It was originally expected that comparisons between different categories of respondents could be made by the use of inferential statistical analysis. However, on inspection of the data it was decided not to, since the sample size is relatively small and many subcategories have a very sparse representation. It was decided that any conclusions from such analysis would not be reliable. The data for analysis are set out in terms of raw scores, percentages, and mean, median, and modal values in this chapter.

4.7 Reliability and Validity

Reliability of the test is deemed to be reflected in the consistency of responses across the group of teachers tested. In the case of the teachers' evaluations, in the present study, it must be noted that the sample number was quite small and this must, of course, be borne in mind when generalising on the findings.

The reliability of the present study is further strengthened by the following:

(a) having an expert group of evaluators

(b) teachers' voluntary participation in the study

(c) the use of Likert scale responses (see Chapter 3.3.4.2, which discusses the reliability of Likert scale questionnaires)

With regard to test validity in the present context, several factors were taken into consideration in the design phase:

Cryer and Home (2008) discuss suggestions that users' acceptance of synthetic speech may depend on the context in which it is being used. Synthetic voices will, therefore, be assessed here also in the context in which they arise (see Chapter 3.3.2.2). For the purposes of the present study, evaluators were first given a sample of the synthetic speech being used as part of the *Digichaint* platform. This was done in order to acclimatise them to the synthetic voices and to the game-like CALL context in which the voices were being used. Respondents then orthographically transcribed each of twenty sentences and gave an immediate reaction on a Likert scale on the 'ease' with which they could transcribe each sentence. The validity of these ratings is considered to be high because of the very short timespan between teachers completing the transcription and rating the 'ease' with which they could do so. This was designed so that one can be assured that the ratings are not influenced by a timespan between performing the transcription and giving an opinion on the 'ease' of transcription. Any such timespan could lead to a certain memory loss and reduce the validity of the subjective response.

It has been claimed that the scores on Likert scales could lack validity because of respondents wishing to give the 'right' answer in the interests of being polite when they are face to face with the person conducting the research (see Chapter 3.3.4.2, (Worcester & Burns, 1975)). In this study, evaluators participated in a relatively anonymous way without being in direct contact with the conductor of the research, and this practice could be expected to increase the validity of teachers' subjective responses (McLeod, 2008).

4.8 Ethics

Ethical approval was sought prior to administering these tests. This involved submitting documentation including a general application form; participant information leaflet; and

the wording of the teacher questionnaire to the School of Linguistic, Speech and Communication Sciences Research Ethics Committee. The project was approved in full.

4.9 Results

4.9.1 Overview.

Results are presented in this section in a format which follows the layout presented in Section 4.3: Research Design & Strategy.

The pre-test questionnaire yielded data which gave a clear description of the background of the teacher respondents (described in Section 4.9.2). The group was predominantly female. They were generally under 40 years old, and had a preference for the Connaught dialect of Irish. They had a general negative predisposition towards synthetic voices.

Results of the objective evaluation (i.e. orthographic transcription task) and the subjective evaluation (i.e. opinions on the ease with which the transcription task could be carried out) are presented under 'Test findings', Section 4.9.3. This section is further divided into two subsections, Sections 4.9.3.1 and 4.9.3.2, which are described briefly here.

Section 4.9.3.1 gives the results of the transcription task (or 'performance' task) and the results of the teachers' opinions on the ease with which they could do the task (or 'opinion' task). The mean score for the subjective evaluation ('opinion' scores) was always considerably lower than was the corresponding score from the objective evaluation ('performance' scores). This shows that performance in itself, as in a transcription task, does not fully reflect the clarity of the synthetic speech, nor the ease with which one can understand it.

In evaluating synthetic speech, one needs to take both factors into consideration – i.e. 'performance' and respondents' perceptions of the ease with which the task may be performed. Previous evaluations of TTS synthesis have relied either on 'opinion' scores (e.g. the ITU-T Recommendations P.85, (1994)) or 'performance' scores (Clark, Podsiadlo, Fraser, Mayo, & King, 2007; Mayo et al., 2011; Valentini-Botinhao, Yamagishi, & King, 2011) none of which are complete in themselves in the context of a general evaluation of CALL platforms incorporating TTS synthesis. In order to compile a scale which takes both factors into account, a formula was developed (see Section 4.3.2.2.2) so as to calculate an index, the *Clarity Index*, which combines 'intelligibility' and 'ease of intelligibility'. The application of this formula and the results obtained are described in Section 4.9.3.2 below.

Results of the short post-test questionnaire are presented in descriptive terms in Section 4.9.4 and a summary of the findings is included in Section 4.9.5.

4.9.2 Teacher pre-test questionnaire findings.

Before taking the test, teachers were asked for some background information. The pre-test questionniare, as presented to the teachers is included in Appendix B1. Of the 31 teachers who participated, 65% were in the under 40 years age group and less than 20% were over 50. Most were reared in non-Gaeltacht areas (N=24), with just 3 respondents coming from the Donegal Gaeltacht and a further 3 from Connemara in the Connaught Gaeltacht. There was 1 respondent from the Gaeltacht in Ring, in Munster. The rest came from a cross-section of counties with Dublin being most strongly represented. The Connaught dialect was the preference for 52% of respondents with only 13% showing a preference for Ulster Irish. The vast majority of respondents reported a 'tolerant' or 'neutral' attitude towards synthetic voices.

The predisposition of the group was, in general, not very favourable towards TTS synthetic voices. 52% could 'tolerate' while 36% declared having a neutral attitude. At

the extreme ends 6% expressed hatred of the voices while, similarly, 6% expressed a positive attitude saying that they are 'sometimes suitable' or 'sometimes more suitable than human voices'. This is broadly in line with many studies which have shown that people prefer human speech to synthetic speech (Stevens, Lees, Vonwiller, & Burnham, 2005), including the Edinburgh University report on 'User Attitudes Towards Real and Synthetic Speech' (CCIR-5, 1999) which found that people had a strong preference for natural speech for an information service. Nevertheless, it shows a preparedness to accept synthetic speech in certain contexts. This same mixed reaction to synthetic speech was reported in a study commissioned by the Centre for Accessible Information of the Royal National Institute of Blind People (Cryer & Home, 2009). The respondents referred to in the Cryer and Home (2009) report were blind and the survey aimed to discover their attitude towards synthetic speech with a view to making more books that are read by synthetic voices available to them. Many respondents reported disliking synthetic speech in general and finding the synthetic voices difficult to listen to, while others saw potential for them in terms of giving access more easily to a wider variety of books.

One may claim then that attitudes of the teachers in this study towards synthetic voices were similar to those reported in Cryer and Home (2009). It may be that this pattern of attitudes towards synthetic voices in general is representative of various sections of the general western population at the present time. The sample size here is such that it was considered unsafe to seek correlations between teachers' predisposition towards synthetic voices and their reactions to the *ABAIR* voices. Nevertheless, the teachers who took part in this study had a high level of literacy and experience of the Irish language (all were Irish language graduates). Therefore, their opinions on the intelligibility and quality of the synthetic speech carry a strong reliability. Table 4.4 below gives a detailed breakdown of the responses to the teacher pre-test questionnaire:

Table 4.4

Teacher pre-test questionnaire results: background information elicited from teachers

			Number	Percentage
1	Total number of participants		31	100%
2	Gender	Female	25	81%
		Male	6	19%
3	Age Group	20 - 30	9	29%
		30 – 40	11	36%
		40 - 50	5	16%
		50 - 60	6	19%
4	Raised in the Gaeltacht	Yes	7	23%
		No	24	77%
4.1	If yes, which Gaeltacht?	Gaoth Dobhair	3	43%
		Conamara	3	43%
		An Rinn	1	14%
4.2	If no, what county were you raised in?	Dublin	11	46%
		Mayo	2	8%
		Limerick	1	4%
		Cork	1	4%
		Wexford	1	4%

		Galway	2	8%
		Carlow	1	4%
		Cavan	1	4%
		Offaly	1	4%
		Meath	2	8%
		Waterford	1	4%
5	With which dialect are you most comfortable?	Ulster	4	13%
		Connaught	16	52%
		Munster	11	35%
6	Describe your attitudes towards synthetic voices	I hate them	2	6%
		Tolerate them but prefer human voices	16	52%
		Neutral	11	36%
		Sometimes suitable	1	3%
		Sometimes more suitable than human voices	1	3%

4.9.3 Test findings.

This Section is divided into two subsections (mirroring the structure of Section 4.3.2.2):

(a) Section 4.9.3.1 describes the 'performance' and 'opinion' results from the orthographic transcription task

(b) Section 4.9.3.2 describes the results of applying the newly developed formula, the *Clarity Index* which combines the 'performance' and 'opinion' (or objective and subjective) scores gleaned from Section 4.9.3.1.

4.9.3.1 Sentence analysis findings.

Detailed results of the analysis of the data for each of the twenty sentences is set out in table format in Appendix D1 and these will be referred to in the discussion below. Figure 4.4 below illustrates in the form of a bar chart one element of the table presented in Appendix D1, namely the modal score results for both the objective and subjective tasks, for each of the sentences in each sentence-category.

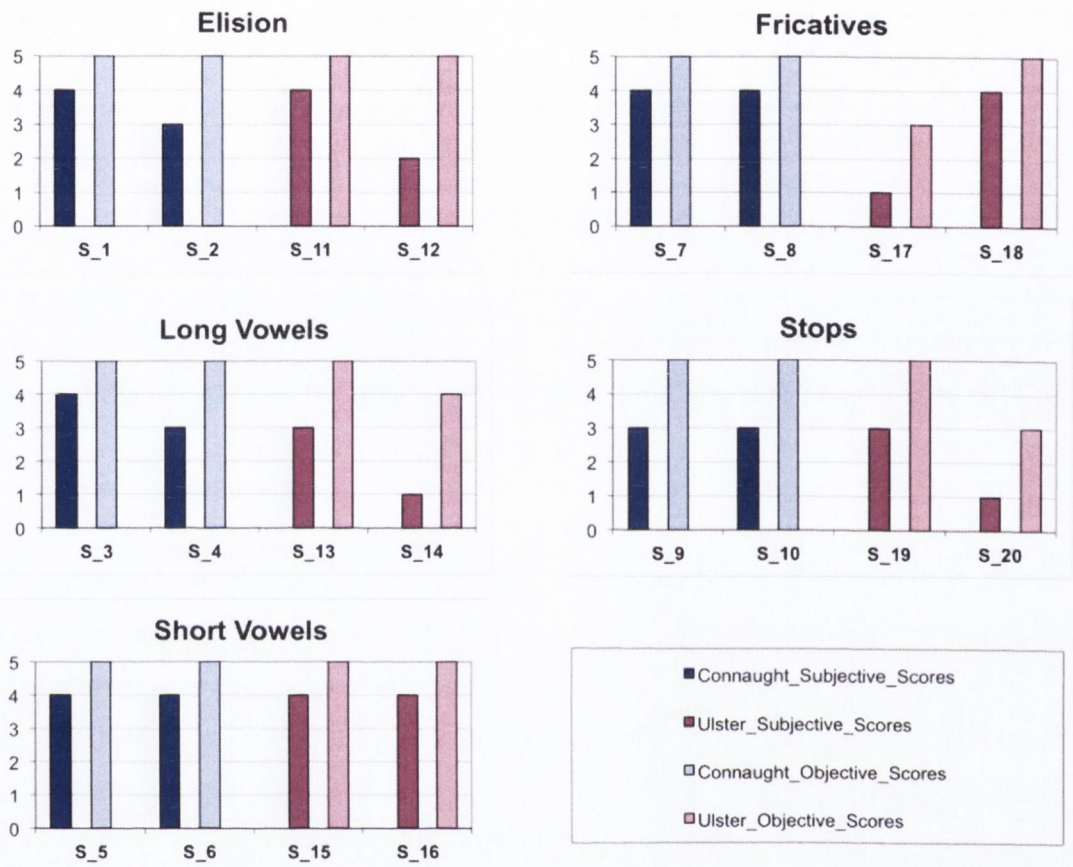


Figure 4.4. Teacher test results: modal values for teachers' results in the 5 sentence categories. To the left of each pair of columns is the opinion ratings the teachers gave in the 'ease' of transcription task and to the right are the performance ratings given to each teacher based on their orthographic transcriptions. Both are marked on a scale of 1 – 5.

The bars on the left hand side of each pair of bars in Figure 4.4 represent the subjective evaluation results and the bars on the right hand side represent the objective evaluation results. The sentences spoken in the Connaught dialect are presented in blue and those in the Ulster dialect are in pink.

One can see at a glance that the mode for the transcription exercise was 5 for all twenty sentences with the exception of three (sentences 14, 17 and 20), whereas the mode for

teachers' rating on the 'ease of transcription' was lower in all cases, with scores ranging from 1 to 4. It may be noted here at the outset that most sentences were intelligible to most respondents, as is evidenced by their high scoring in the transcription (or 'performance') results. It is noteworthy also that the three sentences which did not have a mode of 5 for the transcription exercise involved the Ulster dialect of Donegal.

Overall, scores for ease of comprehension (or 'subjective evaluation results') were low: the mode for teachers' rating on the 'ease of transcription' was lower in all cases, with scores ranging from 1 to 4. Again, it is striking that the lowest scores involve the same sentences (Items 14, 17 & 20) produced with the Donegal voice.

The most difficult item was Item 17 ("*Ar éigean go n-aithneoinn é dá siúlfadh sé isteach an doras anois.*" Translation: '*scarcely would I recognise him if he were to walk in the door now*'). This falls into the 'fricative' sentence-category and was presented with the Ulster unit-selection voice. Its mean score is 3.2 with a median and mode of 3. The respondents' evaluation in terms of the sentences' perceived 'ease of intelligibility' had a mean score of 1.6 but a median and modal score of only 1. This demonstrates that the respondents found that Item 17 was particularly difficult and unclear.

Item 20 ("*Tá turas eagraithe dúinn inniu.*" Translation: *there's a trip organised for us today*) had a somewhat similar scoring pattern. In terms of its objective evaluation it had a mean of 3.6 with median and modal scores of 3. This suggests that while most respondents found the item difficult to understand and transcribe, some still managed a high score on the item. 25% of respondents got the item fully correct.

It is interesting to note the rather large gap between the scores attributed to Item 17 and to Item 18. Both items are classified under the 'Ulster fricative' category. Item 18 had a perfect score of 5 as a mean, median and mode in the objective evaluation whereas Item 17 had much lower scores (mean: 3.2; median: 3; mode: 3). Item 18 also had much higher

scores (mean: 3.7; median: 4; mode: 4) than Item 17 (mean: 1.6; median: 1; mode: 1) in the subjective evaluation, suggesting that teachers did not consider Item 18 particularly difficult to transcribe. The big difference in scores between these two items in the same category could be due to a number of factors, such as the relative complexity of the structures, the relative familiarity of the words/phrases as well as, of course, the fact that there are greater cross dialect differences of pronunciation in Item 17 than in Item 18. Note that the majority of respondents are more comfortable with the Connaught dialect and that the dialectal difference would therefore have a greater impact on intelligibility and ease of transcription for Item 17. IPA transcriptions for both sentences are given in Table 4.5 below, with cross-dialectal differences in bold:

Table 4.5

Dialectal Comparison at a Phonetic Level between Items 17 & 18 of Test

Item 17: 'Ar éigean go n-aithneoinn é dá siúlfadh sé isteach an doras anois.'
Ulster dialect:
e rʲ e: ɟ ə n̥ˠ g ə **n̥ˠ ai ç nʲ o n̥ʲ** e: d̪ˠ a: ʃ u: l̪ˠ **u**: ʃ e: **i** ʃ tʲ a **h** ə n̥ˠ d̪ˠ o ɾˠ ə sˠ ə n̥ˠ i ʃ
Connaught dialect:
ə rˠ e: ɟ ə n̥ˠ g ə **n̥ˠ a n̥ʲ o: n̥ʲ** e: d̪ˠ a: ʃ u: l̪ˠ **ə x** ʃ e: **ə** ʃ tʲ a **x** ə n̥ˠ d̪ˠ o ɾˠ ə sˠ ə n̥ˠ i ʃ
Item 18: 'B'fhearr liom seasamh go fóill, má tá sé sin ceart go leor leatsa
Ulster dialect:
bʲ a: ɾˠ lʲ **o** mˠ ʃ a sˠ **u**: g ə fˠ o: l̪ʲ mˠ a t̪ˠ a: ʃ e: ʃ i nʲ c a ɾˠ t̪ˠ g ə l̪ʲ o: ɾˠ lʲ a t̪ˠ sˠ ə
Connaught dialect:
bʲ a: ɾˠ lʲ **u** mˠ ʃ a sˠ **ə vˠ** g ə fˠ o: l̪ʲ mˠ a: t̪ˠ a: ʃ e: ʃ i nʲ c a ɾˠ t̪ˠ g ə l̪ʲ o: ɾˠ lʲ a t̪ˠ sˠ ə

*Differences between the dialects are denoted in bold font.

Overall, scores for ease of comprehension (or 'subjective evaluation results') were low. One may speculate here that the difficulty may be as much related to the lack of familiarity of the respondents with the Ulster dialect as it has to any inherent problems

associated with the synthetic speech. One may note from Table 4.4 (Item 5: "With which dialect are you most comfortable?") that only 13% of respondents reported being most comfortable with the Ulster dialect. It may very well be the case that those who have a strong familiarity with the Ulster dialect find no particular difficulty with the synthetic speech and that difficulty with the item for others may be due to lack of familiarity with the dialect. One may not make any strong claims in relation to this observation because of the limited number of respondents and the geographical spread in their backgrounds. The question of familiarity with dialects can be further explored in the context of a future systematic evaluation of the synthetic speech output by *ABAIR*.

In general, the mean for each item on the orthographic transcription (or 'objective') task is between 4 and 5, which shows that most respondents found the items intelligible.

4.9.3.2 Clarity Index findings.

The formula, as developed in Section 4.3.2.2.2, is presented here again for convenience of reference. The results obtained from applying this formula to the data collected from 31 respondents, are set out below. Two tables, showing the calculations for the Ulster dialect sentences and the Connaught dialect sentences respectively, are presented in Appendix D2.

The formula,

$$5 - \frac{\sum_{n=1}^{N}(5-P_n)+(5-O_n)}{2n}$$

which can be expanded to,

$$5 - \frac{\big((5-P_1)+(5-O_1)\big)+\big((5-P_2)+(5-O_2)\big)+\cdots\big((5-P_n)+(5-O_n)\big)}{2N}$$

where:

N = total number of respondents

P_1 = the mean performance score of respondent 1

O_1 = the mean opinion score of respondent 1

The mean 'performance' score and 'opinion' score for each of the dialects, based on 20 sentences as judged by the 31 participants, are presented in Table 4.6.

Table 4.6

Mean performance and opinion scores for each of the dialects

	Performance Score	**Opinion Score**	**Clarity Index**
Ulster dialect	4.36	2.86	3.61
Connaught dialect	4.74	3.53	4.14
Average cross-dialect scores	4.55	3.20	3.88

Because of the dialectal variation, separate analysis was done on the Ulster and Connaught dialects. The Connaught dialect got a somewhat higher evaluation on both the performance and opinion scores. The difference between the two may be related to the fact that a majority of respondents reported a preference for Connaught Irish in the pre-

test questionnaire and there were relatively few who expressed a preference for Ulster Irish. This may suggest that the differences in scores may be more closely related to general issues concerning the mutual intelligibility of dialects than to the quality of the synthesised speech output per se for each dialect.

It is difficult to compare performance scores across different studies as the methodologies used for grading the performance have varied somewhat from one case to another. If we combine the mean 'performance' results for both the Ulster and Connaught dialects from Table 4.6 above, we get a score of 4.55, which suggests a 91% accuracy rate. This performance score is very comparable to the results obtained by Al-Wabil, Al-Khalifa and Al-Saleh (2007). They reported a correctness level of 87% to 93% for non-vowelized Arabic words from the different TTS systems and from 89% to 94% for vowelized words. While one cannot draw a direct comparison with the present study because of differences in the scoring methodologies, nevertheless, the broad similarity is worth noting.

If one took these 'performance' scores alone then one would be tempted to conclude that the voices had reached a level of near perfection, similar to what is being claimed for the intelligibility of the better modern TTS synthesisers (Mayo et al., 2011). One must, however, consider the 'opinion' score on the 'ease of intelligibility' of the speech so as to take into account the additional cognitive load associated with processing synthetic speech (van Merrienboer & Sweller, 2005). In the present study, the combined mean 'opinion' score for the two dialects is 3.20, which is very much lower than the 'performance' score, and indicates that the listeners needed to give the utterances greater attention than that which would be necessary in the case of a clear human voice.

An overall *Clarity Index* for the 20 sentences, combining both dialects, as judged by the 31 participants and computed in accordance with the formula presented above is 3.88 (note: this is when cross-dialectal differences are not taken into account). One may argue

that this *Clarity Index* is a fair representation of the 'intelligibility' and 'ease of intelligibility' of the speech samples presented to the participants. A number of studies have evaluated synthetic speech in terms of its "naturalness" (Alvarez & Huckvale, 2002; Lampert, 2004; Sityaev, Knill, & Burrows, 2006) and scores are presented in terms of 'mean opinion scores'. A score of 3.88 is higher than mean opinion scores alone because it also includes a 'performance' score. The present score represents 'clarity' rather than 'naturalness', where clarity here refers to 'intelligibility' and 'ease of intelligibility' combined. This is the most suitable index to report because it best reflects the functionality of the speech in the context of an interactive game-like CALL platform.

The overall value of 3.88 reported in the present study is well below an 'ideal' score of 5. The score suggests that the synthetic speech may be less clear (in the sense intended by the Clarity measure) than natural speech. Our conclusions have to be tentative for a number of reasons. Firstly, we have no available score for natural speech and this is something that future work might address. Furthermore, as was commented earlier, the cross-dialectal difference affected listeners' judgments, particularly in the subject ratings for certain Ulster Irish sentences. All we can say with confidence is that the intelligibility is very high, that there is a dialect effect on 'ease' of intelligibility and that overall, the 'Clarity' measure is quite high. Since perfection eludes even the most advanced TTS systems (and indeed human speech) this score seems quite positive. There are no similar values available from other studies since the combination of 'intelligibility' and 'ease of intelligibility' has not previously been considered in this manner.

4.9.4 Teacher post-test questionnaire findings.

Details of the responses to the questionnaire are set out in Table 4.7 below. These questionnaires were completed when the teachers had finished watching the 3-minute video of *Digichaint* and transcribing and rating the twenty sentences that formed the test. They were asked to pick the statement which most closely reflected their opinion, and the numbers and percentages who chose each of the statements is shown in Table 4.7. Details of the responses obtained from open dialogue boxes, which were also included in the

post-game questionnaire, are not reported here because they did not differ in any significant way from the information gathered through the questionnaires. The teachers tended to be supportive of the concept and expressed good will towards the development of computer-based teaching facilities in the future.

While most (61%) would have a preference for a human voice for a game like *Digichaint*, nobody thought that synthetic voices are 'not suitable at all'. The remainder were more positive towards synthetic voices with almost 40% seeing high potential for the synthetic voice for this type of game.

There was a complete range of opinions on the level of attractiveness of the voices as perceived by the teachers for use with senior pupils in second level schools. Just 9% of them rated the voices 'very attractive' while 13% rated them 'not at all attractive'. The strongest category was at number 4 on the Likert scale, which may be interpreted as 'attractive'. Nobody thought that the quality of the TTS synthetic speech produced by *ABAIR* was 'very weak' and 32% considered it 'excellent'.

One may note the considerable difference between the number of respondents who thought that the voices would be 'very attractive' for pupils (9%) and those who thought that the quality of the synthetic speech was 'excellent' (32%). One possible interpretation of this could be that, while they found the synthetic voices good enough to be understood by themselves, they somehow did not think that pupils would be able to deal with them equally well. It is worth noting in this regard that the pupil responses to the 'attractiveness' of the synthetic voices in *Digichaint* (Item 17_3, Appendix E1) is quite similar to what the teachers thought the attractiveness of the voices would be for pupils. (42% of pupils considered the voices in *Digichaint* to be 'attractive' or 'very attractive' while 51% of teachers thought the voices would be 'attractive' or 'very attractive' for pupils). Table 4.7 sets out the breakdown of the responses to the teacher post-game questionnaire.

Table 4.7

Teacher post-test questionnaire results

			Number	Percentage
Q.	Total number of participants		31	100%
1.	Do you think synthesised speech at this level would be suitable for a game like *Digichaint?*	The synthesised voices are not suitable at all in this game.	0	0%
		A human voice would be preferable to a synthesised voice in this game.	19	61%
		It wouldn't matter whether human or synthetic voices were used.	3	10%
		Synthesised voices are slightly more suitable than human voices for this game.	3	10%
		Synthesised speech would certainly be suitable.	6	19%
2.	Do you think the synthesised speech in the sample you heard would be **attractive** for senior cycle pupils in second level schools? (Please rate on a scale of 1 – 5)	1 = Not at all attractive	4	13%
		2	4	13%
		3	7	23%
		4	13	42%
		5 = Very attractive	3	9%

3.	What is your opinion of the **quality** of the speech produced by *abair.ie*? (Please rate on a scale of 1 – 5)	1 = Very weak	0	0%
		2	1	3%
		3	8	26%
		4	12	39%
		5 = Excellent	10	32%

4.9.5 Summary of findings.

The 20 sentences presented to the teachers were highly intelligible to them and they could orthographically transcribe them with a 91% accuracy rate. This is reflective of other research on the intelligibility of synthetic speech (Al-Wabil et al., 2007). When asked for their opinions on the ease with which they could transcribe each sentence respondents reported some difficulty in so doing. Their performance scores on the transcription test (4.55 (mean performance score)) was higher than their opinion scores on the ease with which they could do so (3.2 (mean opinion score)). The concept of 'clarity' is being introduced and this represents a combination of the performance scores and the opinion scores. Combining these scores has not been done in previous research and it is being argued that it more accurately reflects the quality of the synthetic speech than 'intelligibility' alone. An overall *Clarity Index* for the 20 sentences, combining both dialects, as judged by the 31 participants and computed in accordance with the newly developed formula, is 3.88. One may also note dialectal differences and conclude that the lower scores for the Ulster dialect, particularly in the case of participants' ratings on the 'ease of intelligibility', may be due to a lack of familiarity with the dialect rather than any inherent problem with the synthesis.

Teachers had a mixed reaction to the concept of using synthetic speech as part of a CALL game. Most would prefer a human voice but could accept the synthetic one and just over half of them rated the voices 'attractive' or 'very attractive' and of a very high quality.

4.10 Summary and Relating Teacher Findings to Hypotheses

Chapter 4 has dealt with the reaction of Irish language teachers to the synthetic speech. Based on this research one can get some answers to the overall research questions posed in Chapter 3.4.

It was decided to use teachers for this section of the research as they can be regarded as having a very high level of Irish language comprehension ability. One can reasonably claim that if the synthetic speech is not intelligible to these respondents then the fault lies with the adequacy of the synthetic speech rather than with the inadequacies of the listeners. There is a caveat, however, in that some respondents may find the Ulster dialect unfamiliar.

A distinction was made between the concepts of 'intelligibility' and 'ease of intelligibility'. Respondents were asked to orthographically transcribe 20 sentences which were taken from the text of the *Digichaint* platform and then to rate the ease with which they could transcribe each sentence. The research showed that while respondents could transcribe each with a high level of accuracy, they reported a considerable difficulty in doing so. It has been claimed in the literature that intelligibility of synthetic speech is no longer an issue since it has reached a level of near perfection (Mayo et al., 2011; Nass & Lee, 2001), but these claims are for the best systems for major languages such as English and the intelligibility levels for the newly developed Irish synthesis was not hitherto established. The results from the present research concurs with findings for English synthetic speech in that the performance of respondents in the orthographic transcription test was close to perfect. The respondents reported, however, that they had some difficulty in carrying out the transcription task. It is argued in this study that a more accurate reflection of the quality of synthetic speech is a combination of 'intelligibility' and 'ease of intelligibility', referred to as 'clarity' in this thesis. A formula was devised which gives a *Clarity Index* for the synthetic speech. This index is lower than a

'performance' index alone but may be a more valid representation of the quality of the synthetic speech. A more negative attitude to synthetic speech may be associated with a lack of 'ease of intelligibility'. The cognitive processing load associated with synthetic speech would appear to be greater than that associated with natural speech (see Chapter 2.5.6.1).

The majority of respondents in this section of the study were female, under forty years of age, and they had a generally negative predisposition towards synthetic speech. Results from the research with this group lends weight to some of the hypotheses pertaining to the main research questions as stated in Chapter 3.4.

Hypothesis 2 states that "the synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability". One can claim with a high level of certainty, that the utterances are largely intelligible to the teacher respondents, which supports this hypothesis. The modal result for the transcription of practically all the sentences was 5, or maximum score. It must also be borne in mind that those sentences were presented in random order and without a context which, one would expect, would make them a little more difficult to comprehend. It was also anticipated that these sentences would be problematic (as discussed in Section 4.3.2.2.1) and a set of sentences randomly selected from the *Digichaint* corpus could be expected to yield a higher score. Respondents managed to achieve this result with relative ease as shown by their subjective scores. A modal score of '4' for 'ease of intelligibility' for 10 of the 20 test sentences is quite high.

Hypothesis 3 states that "the synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness". Teachers' opinions of the quality of the synthetic voices were positive, as can be gleaned from Items 2 and 3 in Table 4.7 above. These results lend support to the hypothesis, insofar as the quality of the synthetic speech was concerned. Nobody thought that the quality was 'very weak' and over 70% gave it a point 4 or point 5 ranking on the Likert scale. They were not, however, as positive about

the 'attractiveness' of the voices, with only a small majority (51%) rating them as being 'attractive' or 'very attractive'.

Hypothesis 4 states that "a majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented". Data from the teachers' responses (Table 4.7) showed that 71% of respondents rated the quality of the voices at points 4 or 5 on the Likert scale (point 5 described as 'excellent'). Nobody thought that the voices were 'not suitable' for a game like *Digichaint* (Item 1). 51% of the teachers rated the voices at 4 or 5 on the Likert scale in their opinions of the attractiveness of the voices to pupils. Taken together these points offer support for the hypothesis.

This study is not a general evaluation of the speech produced by *ABAIR*. It is attempting to establish if the synthetic speech is of a sufficiently high standard to be used in a game-like CALL platform. Based on the present evidence from the teachers' responses, the answer is positive. It was not considered useful to further analyse the group so as to establish which characteristics of respondents were most related to their response patterns. This may be done in a future study which evaluates the overall output of the ABAIR synthesis and has a much higher number of evaluators. Overall, numbers were relatively small and the group relatively homogenous, rendering statistical analysis unsafe. One could not claim with any reasonable degree of certainty that this group is representative of the entire population of those with a high standard of Irish. Analysis of the data was therefore not carried to an inferential level.

Chapter 5: Findings for Learners

5.1 Overview

This chapter focuses on the pupils' evaluations of the three platforms. It gives detailed evaluations of the platforms in terms of their general design and functionality (non-TTS aspects) and also gives the pupils' evaluations of the synthetic speech (TTS) which is integrated into the newly developed game-like CALL platforms (see Chapter 3.2).

Unlike the evaluations by teachers, described in Chapter 4, present evaluations are almost entirely questionnaire-based and represent the opinions of the pupil sample. The rationale for using pupils as evaluators in the study is discussed in Chapter 3.3.3. It was decided to avoid the orthographic transcription tests, similar to those used with the teachers, as the results of such tests may lack reliability. Deficiencies in the results could be ascribed to any one or a combination of a number of issues:

(a) poor quality synthetic voices

(b) low comprehension skills of the pupils

(c) poor orthographic skills of the pupils

The main aim of this part of the study is to test each of the hypotheses stated in Chapter 3.4 using the subjective data acquired from the pupils' responses to questionnaire items relating to each of the three platforms which have integrated synthetic speech. It sets out to get pupils' reactions to both linguistic (TTS) and non-linguistic (non-TTS) aspects of the platforms.

This chapter begins by giving background information on the participants and the reasoning behind the choice of schools taken for the sample. Section 5.2 describes the

participants who took part. These were mainly 17-year-old pupils from three distinctive linguistic backgrounds - pupils from Gaeltacht schools, Gaelscoileanna (Irish-medium schools outside the Gaeltacht) and English-medium schools. In total, 13 schools were identified as having pupils who formed a representative sample of the type of pupils required for the evaluations.

A description of the design of the study is set out in Section 5.3 of this chapter. Pupils (N=252) were introduced to each of the three platforms and asked to participate in the activities. They were first given a 'pre-game' questionnaire in order to elicit some background details pertaining to each individual respondent. They then played the 'game' and, finally, gave their reactions to each platform by way of a 'post-game' Likert scale-based questionnaire. (The term 'game' is being used here to describe collectively the activities involved in each platform. Strictly speaking only one platform, *Digichaint*, has a game-like structure, as described in Chapter 3.2).

Details are given on the instruments used to gather the data in Section 5.4 and on how each item in the questionnaires relates to specific points raised in the theoretical background to the thesis (see Chapters 2 and 3).

Section 5.5 below sets out the procedures used when collecting the data and Section 5.6 deals with the system of analysis used. Issues pertaining to validity and reliability are discussed within Section 5.7 and the procedures involved in getting ethical approval for the project are mentioned in Section 5.8.

Section 5.9 presents the results. Two types of analysis used. Firstly, results of the 'pre-game' questionnaire and 'post-game' questionnaire are presented in terms of raw scores or percentages which are referred to as descriptive statistics. In the second part of the analysis, correlations between the data obtained from the 'pre-game' questionnaire items and the data obtained from the 'post-game' questionnaire items are calculated using the

Kruskal-Wallis technique. This is used to test for relationships between the five background factors selected (namely, 'gender'; 'school type'; 'frequency of playing computer games'; 'level of understanding of Irish' and 'attitude towards synthetic voices') and respondents' opinions of the various facets of the platforms such as graphics, plot, intelligibility of synthetic speech, etc. Finally, a Spearman's rho correlation technique is used which tests the correlation between the data elicited from all items in the 'post-game' questionnaire and tests for significant relationships. This was done in order to test the proposition that engaging with each of the platforms is a holistic experience and that one's opinion of any aspect of the platform will influence and be influenced by one's attitude to all the other aspects of the platform, including the synthetic speech.

Results of the analysis are set out separately for each of the three platforms under Section 5.9.3 and a summary of these inferential statistical analysis is given in Section 5.9.3.3.5.

Section 5.9.4 discusses the implications of the results for each of the hypotheses in turn and finds that the data supports them. There is some discussion on the main findings in the final section of this chapter in Section 5.9.5.

5.2 Participants

Evaluators of TTS synthesis tend to be drawn from native or near-native speakers of the language being synthesised. It is also suggested that evaluations should be carried out in situations which are similar to those in which applications incorporating the synthetic speech would be used in the future (see Chapter 3.3.2).

In the case of TTS synthesis being used as a facet of a CALL platform, the end users are language learners, and although some studies have used learners as evaluators of TTS synthesis (Kang et al., 2008), the use of learners gives rise to a potential problem. When

using learners poor evaluation results could be due to a number of factors: inadequacies in the synthetic voices; issues pertaining to the CALL platforms; or they could also quite simply be due to the language level of the evaluators.

The present study is evaluating three CALL platforms with integrated TTS synthesis. As mentioned earlier, evaluation of the TTS synthetic voices and the other non-TTS dimensions of the platforms, though separate, are likely to be related. It is to be expected, and it is a specific hypothesis in this thesis (see Hypothesis 5, Chapter 3.4), that these aspects will influence each other. As the learner's language level could influence both aspects of evaluation, it was felt important that, at least for some of the evaluators, the further variable of the language learner's competence be eliminated, or at least reduced, by ensuring a strong cohort of 'competent' pupils.

The inclusion of teachers as the first evaluator group was also one way of mitigating the risk of results being skewed by the language level of the evaluators. And, as was explained in Chapter 4, the teachers were also deemed to be the most appropriate group from which to elicit intelligibility ratings. As regards the pupil evaluators, a group of 16-17 year olds (N=252) was selected, as this is the age group for which the platforms are developed. It was felt important that two subgroups of evaluators be included. First of all, one needed to test pupils at whom these platforms are targetted (i.e. pupils in English-medium schools, whose command of Irish may be rather poor). Secondly, it was felt important to have a substantial number of evaluators whose competence in Irish would be at a high level so that language competence would not impact negatively on the evaluations of the platforms and the synthetic voices. This latter group is drawn from two backgrounds, those from the Gaeltacht and those from Gaelscoileanna (Irish-medium schools). The complexity of the socio-cultural position of Irish is set out in Chapter 2.2 above. There is a great deal of variation, in terms of Irish language ability, amongst pupils from different sociolinguistic backgrounds. To summarise, the pupil evaluators were chosen from the following three categories:

(1) English-medium schools

(2) Gaelscoileanna (Irish-medium schools situated in English speaking parts of the country)

(3) Gaeltacht schools (Irish medium schools situated in Irish speaking regions)

These broad categories are similar to those used by Harris, Forde, Archer, Nic Fhearaile, and O'Gorman (2006) in their study which was a national assessment of achievement in the Irish language of sixth grade primary school pupils with regard to their listening, speaking and reading levels. A surprise finding in that study was that, although intuitively one would expect that pupils from Gaeltacht areas should have the highest language levels, the highest performances came not from Gaeltacht schools but rather from Gaelscoileanna. It should be noted that both Gaelscoileanna and Gaeltacht pupils were significantly ahead of those from English-medium schools in their levels of achievement. There are many complicating factors which may be associated with the results of this study - for example, the social and economic backgrounds of the pupils' parents, and the linguistic variation in the Gaeltacht where increasingly the numbers of non-Irish speakers are encroaching (see discussion in Chapter 2.2). When it comes to comparisons between Gaeltacht native speakers and Irish speakers from a Gaelscoil, one has to be cautious in drawing conclusions from a study such as the above, since the Irish of native speakers from Gaeltacht areas tends to be very localised and quite different from the type of standardised language which is used for national standardised tests (Ó Flatharta, 1989).

Returning to the present study, the pupil sample selected is not intended to be a representative sample of pupils nationally. Proportional to the national population, it has an overrepresentation of the highly proficient group, particularly those from Gaeltacht schools, as shown in Table 5.1 below. As mentioned, the assumption is that Gaeltacht and Gaelscoil pupils generally have a higher standard of Irish than those in English-medium schools. This is supported by Harris et al. (2006) and corresponds to my own intuition and personal experience. A further assumption is entailed here - that the general Irish

language ability of Gaeltacht and Gaelscoil pupils is beyond that required for the level of language presented in the present CALL platforms, given that the language level was specifically designed for those from English-medium schools.

Although the initial expectation is that Gaelscoileanna and Gaeltacht schools broadly form a coherent grouping of 'competent' pupils in regard to Irish, the analysis treats all three school-types as independent variables. Note that there no studies which compare the Irish language comprehension ability of pupils from each of the three school types in the *second level* school sector have previously been conducted. It has been hypothesised (Hypothesis 2, Chapter 3.4) that one's judgment on the intelligibility, quality and attractiveness of the synthetic speech will be directly related to one's general Irish language comprehension ability. In the statistical analysis of the data from this section of the study, 'school-type' is regarded as an independent variable and is used as a proxy for language ability level.

In relation to the non-TTS aspects of these platforms (judgments on the graphics, plot, etc.), there is no apriori assumption that school-type (proxy for language proficiency) will cause systematic differences, except insofar as the holistic experience can have a knock-on effect. Between-school-differences on both TTS- and non-TTS-related dimensions of the evaluation are tested and results are reported in respect of pupils' reactions to the various aspects of each of the three platforms separately in the section dealing with 'Post-game inferential statistical results', Section 5.9.3.3.

Between-school-type differences can be thought of as being primarily linguistic (i.e. judgments on the intelligibility, quality and attractiveness of the synthetic voices). The opinions of respondents on other aspects of the platforms (e.g. graphics; plot; etc.) could not be expected to systematically vary from one school type to another as significant differences between students' experiences with digital media are unlikely.

Table 5.1

Comparison between distribution of school-types nationally and distribution of school-types and participants selected for the present study

	Numbers Nationally (2013)	**Sample for Present Study – Schools:**	**Sample for Present Study – Participants:**
English-medium schools	655 of 721 schools (91%)	8 of 13 schools (61%)	144 pupils (57%)
Gaelscoileanna (Irish –medium schools)	36 of 721 schools (5%)	1 of 13 schools (8%)	23 pupils (9%)
Gaeltacht schools	30 of 721 schools (4%)	4 of 13 schools (31%)	85 pupils (34%)

In total, 13 schools were identified as having pupils to represent the school-going population which was required for this study. The participants selected comprised 4th and 5th year pupils, generally in the 16-17 year old age-bracket. It was intended that all sections of the Irish school community would be represented in the study, and a good geographical spread was included incorporating schools from each of the four provinces and a mixture of urban and rural settings. In making this selection, one could be reasonably confident that one was including a good spread of Irish language abilities, and therefore one may test for the significance of the relationship between the language competence of the respondents and their opinions on the three platforms being studied. The inclusion of the different categories of students allows for the possibility of inter-group comparisons which could have implications for the conclusions that might be drawn from this study. Table 5.2 below shows the numbers from each of the school types selected for participation in the study. A more detailed breakdown of participants in terms

of gender; attitude towards synthetic voices; self-evaluation of comprehension competence; frequency playing computer games; etc. was gleaned from a 'pre-game' questionnaire and is reported in the findings in Section 5.9.2 below.

Table 5.2

School-types and details on participants selected for the present study: a tabular breakdown of schools and students according to the relevant socio-demographic and sociolinguistic variables

School	Location	School Type	Dialect Preference	Gender	Total Number in Sample	% of Total Sample
1	Rural	Gaeltacht	Connaught	Mixed	24	9.52%
2	Rural	Gaeltacht	Ulster	Mixed	22	8.73%
3	Rural	Gaeltacht	Munster	Mixed	17	6.74%
4	Rural	Gaeltacht	Munster	Mixed	22	8.73%
5	Urban	Gaelscoil	without preference	Mixed	23	9.13%
6	Rural	English-medium	without preference	Mixed	15	5.95%
7	Urban	English-medium	without preference	All girls	9	3.57%
8	Urban	English-medium	Ulster	All girls	15	5.95%
9	Urban (majority)	English-medium	without preference	Mixed	7	2.77%

School	Location	Type of School	Dialect Preference	Gender	Total Number in Sample	% of Total Sample
10	Urban	English-medium	without preference	All girls	19	7.54%
11	Urban	English-medium	without preference	All girls	24	9.52%
12	Rural	English-medium	Connaught	Mixed	27	10.71%
13	Urban (majority)	English-medium	Munster	All girls	28	11.11%

5.3 Learners: Research Design and Strategy

5.3.1 Research Design.

The study design is intended to assess the overall quality of the three game-like platforms, including the non-linguistic aspects (i.e. graphics, plot, etc.) as well as the synthetic speech.

This learner section of the research involves a 3-stage multi-method design (Creswell, 2009) comprised of three instruments: a pre-game questionnaire, a game and a post-game questionnaire.

Stage 1	Stage 2	Stage 3
Pre-game Questionnaire → Elicits background information from the participants	**Game** → Participants play the game	**Post-game Questionnaire** Elicits opinions about game features including plot, graphics, synthetic voices, etc.

Figure 5.1. Research Design for Learner Evaluation

The order in which the instruments were presented to the participants is shown in the three stages in Figure 5.1. This order was really only necessary to answer one subsection of the study: Hypothesis 6 (see Section 3.4), which investigated whether there was a relationship between respondents' predisposition towards synthetic voices and their attitude towards the Irish synthetic voices presented in the game.

This design allowed testing for relationships between Stage 3 data and background variables gathered in the pre-game questionnaire (Stage 1 in Figure 5.1 above). It also allowed for the testing of relationships between participants' responses to the various post-game questionnaire items elicited at Stage 3.

5.3.2 Research Strategy.

Since both teachers and second level pupils had been chosen as evaluators in the overall research and since it was decided to target a relatively large number of second level pupils, it was determined that the use of questionnaires was the most sensible option for data collection. There are several advantages to using questionnaires for the purposes of the present study. The most evident of these is practicality. Using questionnaires, structured data can be collected from a large number of respondents in a relatively short period of time. As the learner evaluation was to be carried out in second level schools, the

questionnaire had to be designed in such a way as to take account of the time limitations imposed by class periods. The use of relatively short, focused questionnaires limits the potential for 'fatigue effects' (Dörnyei, 2003) having a negative impact on pupils' responses to the questionnaire items. Questionnaires allow for eliciting pupils' opinions on items directly related to the TTS and non-TTS aspects of the research questions in a concise format.

There are, of course, potential disadvantages to the use of questionnaires in that questions can be raised as to how truthful a respondent is being; or how much thought a respondent is giving to each answer; or respondents opting to give the 'polite' answer (see Chapter 3.3.4.2). As noted in the 'Procedure' section below (see Section 5.5), I was present at each of the evaluations. I could, therefore, take note of any obvious instances of pupils losing interest (by noting their finish times as being very different from those of the rest of the group, for example, and by scanning the results afterwards to ensure a reasonable spread of responses). An open-ended dialogue box was also included in the design of the questionnaire so that pupils were free to leave either positive or negative comments on any aspect of the study.

The questionnaires were conducted by way of Likert 5-point scales, as mentioned in Chapter 3.3.4.2. It was decided to present the questionnaires to the evaluators in the form of a web-based survey. This involved designing and creating a webpage with the content of the questionnaires using HTML and writing a PHP script to collect the form data the participants submitted and send it to a database. It was important to plan this step carefully and structure it properly from the start so as to ensure the data were organised in such a way that tabulation would be relatively straightforward once I had finished collecting the data. Dörnyei and Tatsuya (2010) mention that respondents are prone to leaving out some questions, either by mistake or because they did not like them. The design of the present study ensured that online questionnaires were set up such that respondents could not progress to a subsequent question without having answered all

previous questions. There were, therefore, no gaps in the responses to individual questionnaire items. Each pupil was given an individual user ID so that his/her personal identification would be protected when the responses were being processed and categorised for results analysis. Both an English and an Irish language version of the questionnaire was made available to all respondents. Irish was set as the default but pupils were free to use the translated version which was available by clicking the 'English' tab at the top of the questionnaire. This was included in order to ensure that lack of Irish comprehension would not cause pupils to give a response without having fully understood the question. This webpage was hosted on and run from the *ABAIR* server.

The information obtained from students in this section of the study is entirely subjective, in the sense that questionnaires are used to elicit pupils' *opinions* of various aspects of platforms, and no attempt is made to elicit any type of objective information such as that obtained in the teachers' section of the study. It was decided that an orthographic transcription exercise, such as that presented to the teachers, would be inappropriate for a number of reasons. One reason relevant here is that it would take from the overall gaming experience of the platforms which was also being enquired into. It could turn the entire exercise into a rather dull experience and have an unduly negative influence on pupils' opinions of the platforms in general.

Background information on the respondents was elicited by way of a 'pre-game' questionnaire. Details of the pre-game questionnaire are set out in Section 5.3.2.1 below.

Participants were then asked to interact with each of the platforms, as described in Chapter 3.2. Immediately after each game, they were asked to complete a post-game Likert scale-based questionnaire which elicited their attitudes to the various aspects of the games being inquired into. Details of the post-game questionnaire are given in Section 5.3.2.2 below.

5.3.2.1. Pre-game questionnaire.

A questionnaire to elicit background information in relation to each of the respondents was designed. This is being referred to here as a 'pre-game questionnaire' since it was designed to be completed by respondents before they had any experience of any of the platforms. It was intended for use as a basis for discovering if these background factors influence the pupils' responses to questions which were administered after exposure to each of the 3 platforms, i.e. their attitude towards the different aspects of the platforms, such as graphics, TTS synthetic voices, etc. The questionnaire administered after the pupils had exposure to the platforms is referred to as a 'post-game questionnaire'.

The pre-game questionnaire items elicit the following information:

1. gender
2. pupils' year in school
3. pupils' school type (Gaeltacht/Gaelscoil/English-medium school)
4. the frequency with which pupils play computer games generally
5. pupils' perception of their own ability level at Irish
6. pupils' rating of their 'Top 3' favourite subjects
7. pupils' preconceived attitude towards TTS synthetic voices in general

Some of the items in the pre-game questionnaire were prompted by the review of the literature (points 4 and 7 below arise as possible salient factors from the review of the literature i.e. pupils' attitudes towards synthetic voices in general (prior to hearing the *ABAIR* voices) (see Chapter 3.3.2.1) and the frequency with which they play computer games (see Chapter 2.5). Others, such as gender and pupils' year in school are included, as intuitively one feels that there may be a gender and age difference in pupils' reactions to online environments. Pupils' school type and their perceived ability level at Irish were intended to elicit information on their Irish language comprehension ability level, as these

could have major implications for their results. Findings from the pupil pre-game questionnaire are reported below in Section 5.9.2.

5.3.2.2. Post-game questionnaire.

As described above, the post-game questionnaire is designed to elicit opinions from students with regard to various aspects of the platforms. A distinction is being made here between information sought on the intelligibility, quality and attractiveness of the TTS synthetic voices (described in Section 5.3.2.2.1) and opinions on the other general characteristics of each of the platforms being evaluated (described in Section 5.3.2.2.2). The actual post-game questionnaires in full, as presented to the pupils, are available in both English and Irish in Appendices C2 & E1 *(Digichaint)*, C3 & E2 (*Fáilte go TCD*) and C4 & E3 (*Taidhgín*).

5.3.2.2.1 Eliciting pupils' opinions on the TTS voices.

The following is a summary of the questions that were used in order to elicit information from the pupils in respect of the intelligibility, quality and attractiveness of the TTS synthetic voices.

1. Respondents were asked for their opinion on whether the language was at a suitable level of difficulty. If they responded that it was not at a suitable level of difficulty, they were further asked to state whether they found that the difficulty level was too high or too low. This item was included because if the level of the language was too difficult for the learners one could anticipate they would then give a lower rating to the intelligibility of the voices, not necessarily because of any problem with the quality of the synthetic voice but rather because of their lack of understanding of the language.
2. They were asked whether they thought the synthetic voices were sufficiently clear to make the speech intelligible.
3. Their opinion was sought on whether the dialects used posed particular problems for their understanding of the synthetic speech.

4. They were asked for a rating on the attractiveness of the voices, where *attractiveness* was left to their own interpretation of that word.

5. They were also asked for their opinion of the overall quality of the voices, and their adequacy, for these game-like platforms.

Because *Digichaint* has stronger gaming features, pupils were asked for the following additional information:

6. to give their opinion on the idea that synthetic speech can give a better atmosphere to the game than can natural voices

7. to compare their level of understanding of natural voices to that of the synthetic voices used in the game

8. to give their opinion on the suitability of the synthetic voices for the particular *Digichaint* game

5.3.2.2.2 Eliciting pupils' opinions on the CALL materials.

The items on the questionnaire which refer to the non-linguistic (non-TTS) aspects of the platforms are summarised here. Students were asked their opinions on:

- the graphics
- the usefulness of the game/material for language learning
- the storyline used in this game. *Note*: *this is only relevant to the Digichaint platform since plot had a central role in this application, unlike the other two*
- the concept of games (generally) for language learning

5.4 Instruments

This section demonstrates how the particular questionnaire items, which are included in the post-game questionnaire for each of the three platforms, were arrived at. As is clear

from the Design section above, some of the items are eliciting opinions to do with the synthetic voices (TTS), whereas others are eliciting opinions on various other (non-TTS) aspects of the platforms. The items for each of the three platforms are set out below in table format (Tables 5.4 (*Digichaint*), 5.5 (*Fáilte go TCD* and *Taidhgín*). The questionnaires are presented in full in Appendices C 2 - 4 (in Irish) and E (English translation).

To facilitate the subsequent discussion, each questionnaire item is presented in Tables 5.4 and 5.5 (rightmost column) and is explicitly linked to the theoretical background/gaming criteria on which it is based (middle column) and, lastly, to the hypothesis it seeks to test (leftmost column). Also, it was felt helpful to include here in Table 5.3 a summary of the hypotheses of the study, which were presented earlier in Chapter 3.4.

Table 5.3

Overview of Hypotheses, as presented first in Chapter 3.4

Hypothesis 1	*Digichaint,* as an interactive game-like platform, will conform to the **criteria** for effective CALL platforms:
1.1	game is task-based with a focus on meaning
1.2	game has an appropriate language difficulty level
1.3	game is easy to manoeuvre
1.4	game is enjoyable
1.5	game has a clear plot
1.6	game has a logical structure
1.7	game is engaging for the learners
1.8	game has graphics that are considered attractive/engaging

1.9	game has legitimate narrative, i.e. credibility for the player/learner
Hypothesis 2	The synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability.
Hypothesis 3	The synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness.
Hypothesis 4	A majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented.
Hypothesis 5	There is a statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms (such as plot, ease of manoeuvring characters, legitimacy of narrative and standard of Irish language used) and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform.
Hypothesis 6	There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general and their attitude towards the Irish synthetic voices being tested.

For the non-TTS aspects of the platforms, questionnaire items (which fall under Hypothesis 1) are based on the 'Criteria for Evaluating the Present CALL Platforms' in Chapter 3.3.1.2 (Table 3.2) which were established on the basis of current theory. There is a column entitled 'Theoretical Background' in Table 3.2 which references the background literature review from which these criteria were drawn. The specific research questions being raised which are essentially the subsections of Hypothesis 1, and the questionnaire items in turn, were derived from these criteria. To recap on the theoretical background, the reader is referred to Table 3.2 and its referenced sections in Chapters 2 and 3.

For the TTS aspects of the CALL platforms, the relevant sections from the literature review are referenced directly in the tables below. There is a certain amount of crossover where one questionnaire item might fit more than one research question or vice versa.

Table 5.4

Hypotheses, Theoretical Background, and their relation to Questionnaire Items for Digichaint.

Hypotheses		Theoretical Background / CALL-based Criteria	Questionnaire Item
Hypothesis 1		**CALL-based gaming criteria formed on the basis of theoretical background and summarised in Table 3.2**	**Questionnaire Items**
1.1	*- game is task-based with a focus on meaning*	**Criterion 1:** Games to be task-based with focus on meaning and on use of target language.	**Item 9:** I was more focused on the plot of the game than I was on the language being used **Item 10:** There is a good balance between enjoyment and language learning in this game
1.2	*- game has an appropriate language difficulty level*	**Criterion 2:** Learners' opportunity to develop meta-linguistic awareness. Language difficulty level. Appropriateness of manoeuvring through games.	**Item 11:** The overall standard of the Irish used in this game is at about the right level for me **Item 3:** I learned some new phrases/words/grammar points as I played the game. **Item 2:** Did you feel the game was technically very difficult to play – very easy to play?
1.3	*- game is easy to manoeuvre*	**Criterion 5:** ...Appropriate clues to successfully manoeuvre through game. **Criterion 9:** Clarity of screen layout.	**Item 2:** Did you feel the game was technically very difficult to play – very easy to play? **Item 1:** Rate the graphics used in this game on a scale of 1 – 5

Hypothesis 1 (continued)		CALL-based gaming criteria formed on the basis of theoretical background and summarised in Table 3.2	Questionnaire Items
1.4	*- game is enjoyable*	**Criterion 3:** Playful spirit of game. Enjoyment level of game.	**Item 5:** Rate your overall enjoyment of this particular game as a language learning experience **Item 10:** There is a good balance between enjoyment and language learning in this game **Item 17_1:** Give your opinion on the usefulness of the concept of producing an interactive language learning game in order to practise Irish
1.5	*- game has a clear plot*	**Criterion 4:** Clarity of game plot. Helpfulness of hypertext dictionary.	**Item 6:** The plot of the game was clear to me from the outset. **Item 4:** The dictionary is helpful as an aid to learning Irish in this game
1.6	*- game has a logical structure*	**Criterion 5:** Logical structure of game. Appropriate clues to successfully manoeuvre through game.	**Item 5:** Rate your overall enjoyment of this particular game as a language learning experience. **Item 6:** The plot of the game was clear to me from the outset.
1.7	*- game is engaging for the learners*	**Criterion 6:** Intensity of engagement with game.	**Item 8:** The game held my attention
1.8	*- game has graphics that are considered attractive/engaging*	**Criterion 7:** Appropriateness of visual/aural material in the game	**Item 1:** Rate the graphics used in this game on a scale of 1 – 5
1.9	*- game has legitimate narrative, i.e.*	**Criterion 8:** Legitimacy of game	**Item 7:** Was the plot credible for this type of game?

	credibility for the player/learner	narrative.	**Item 11:** The overall standard of the Irish used in this game is at about the right level for me.
Hypothesis 2		**Theoretical Background**	**Questionnaire Items**
The synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability.		Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	Background factors X **Item 12:** The synthesised voices were sufficiently clear to make the speech intelligible. **Item 14:** I found it no more difficult to understand the computer-generated voice than I would if natural voices were used. **Item 15:** In general, I think that computer-generated voices give a better atmosphere to computer games than natural voices do. **Item 16:** I found the computer-generated voice suitable for this computer game **Item 17_2:** Give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of game presented here? **Item 13:** Did you experience particular difficulties with the dialects that are used in Digichaint?
Hypothesis 3		**Theoretical Background**	**Questionnaire Items**
The synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness.		Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	**Item 17_2:** Give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of game presented here?

		Item 17_3: Give your opinion on the attractiveness of the voices.
Hypothesis 4	**Theoretical Background**	**Questionnaire Items**
A majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented.	Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	**Item 16:** I found the computer-generated voice suitable for this computer game **Item 17_2:** Give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of game presented here?
Hypothesis 5	**Theoretical Background**	**Questionnaire Items**
There is a statistically significant relationship between the measured reactions of the respondents to the non-TTS elements of the platform and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform.	See discussion in Chapter 2.5.6, in particular" "...game playing is a holistic experience" "...one's view of particular aspects of the game, for example graphics, plot (or TTS voices) is influenced by one's disposition towards the whole of the game..."	All post-game questionnaire items
Hypothesis 6	**Theoretical Background**	Questionnaire Items
There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general & their attitude towards the Irish synthetic voices being tested.	See discussion in Chapter 3.3.2.1.1 and "...evidence to suggest that one's prior experience of synthetic voices may have an influence on one's attitude towards a synthetic voice in a particular context (Cryer & Home, 2009)"	Background factors X **Item 12:** The synthesised voices were sufficiently clear to make the speech intelligible. **Item 14:** I found it no more difficult to understand the computer-generated voice than I would if natural voices were used. **Item 15:** In general, I think that computer-generated voices give a better

		atmosphere to computer games than natural voices do. **Item 16:** I found the computer-generated voice suitable for this computer game **Item 17_2:** Give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of game presented here? **Item 17_3:** Give your opinion on the attractiveness of the voices.

The following Table 5.5 brings together the questionnaire items used in both *Fáilte go TCD* and *Taidhgín*. The questionnaire items used for these two platforms are not identical but are very similar and can be summarised here together. The table follows the format of Table 5.4 (*Digichaint*) but since neither *Fáilte go TCD* nor *Taidhgín* conform fully to the task-based gaming criteria, some of the elements which were used in the evaluation of *Digichaint* are no longer relevant.

Table 5.5

Hypotheses, Theoretical Background and their relation to Questionnaire Items for Fáilte go TCD and for Taidhgín

Hypotheses		Theoretical Background / CALL-based Criteria	Questionnaire Item
Hypothesis 1		**CALL-based gaming criteria formed on the basis of theoretical background and summarised in Table 3.2**	**Questionnaire Items**
1.1	*- game is task-based with a focus on meaning*	**Criterion 1:** Games to be task-based with focus on meaning and on use of target language.	N/A
1.2	*- game has an appropriate language difficulty level*	**Criterion 2:** Learners' opportunity to develop meta-linguistic awareness. Language difficulty level. Appropriateness of manoeuvring through games.	**Item 8:** The overall standard of the Irish used is at about the right level for me.
1.3	*- game is easy to manoeuvre*	**Criterion 5:** ...Appropriate clues to successfully manoeuvre through game. **Criterion 9:** Clarity of screen layout.	N/A
1.4	*- game is enjoyable*	**Criterion 3:** Playful spirit of game. Enjoyment level of game.	**Item 2 (*Taidhgín*):** Do you feel the talking monkey adds to the overall playfulness of the exercise? **Item 4 (*Taidhgín*):** To what extent do you think this type of learning platform (i.e. the interaction and playfulness)

			would help in prasticing conversational Irish? **Item 5:** Would you enjoy using this type of activity to develop your aural Irish skills, should it be available and easily accessible in your school? **Item 6:** How motivating do you find this type of activity? **Item 7:** Do you think this type of activity would make the learning of Irish more attractive?
1.5	*- game has a clear plot*	**Criterion 4:** Clarity of game plot. Helpfulness of hypertext dictionary.	N/A
1.6	*- game has a logical structure*	**Criterion 5:** Logical structure of game. Appropriate clues to successfully manoeuvre through game.	N/A
1.7	*- game is engaging for the learners*	**Criterion 6:** Intensity of engagement with game.	**Item 6:** How motivating do you find this type of activity? **Item 7:** Do you think this type of activity would make the learning of Irish more attractive? **Item 4:** To what extent do you think this type of learning platform (the voices, the graphics and the setting) would help in practising listening comprehension? **Item 11_1:** Please give your opinion on the usefulness of producing graphics with synthesised voices in order to practice aural comprehension.

1.8	*- game has graphics that are considered attractive/engaging*	**Criterion 7:** Appropriateness of visual/aural material in the game	**Item 1:** How would you describe your judgment of the background setting and the graphics in the video? **Item 2:** How would you describe your judgment of the body movements of the figures and their alignment to speech? **Item 3:** To what extent do the movements of the characters add credibility and clarity to the conversational exchanges?
1.9	*- game has legitimate narrative, i.e. credibility for the player/learner*	**Criterion 8:** Legitimacy of game narrative.	N/A
Hypothesis 2		**Theoretical Background**	**Questionnaire Items**
The synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability.		Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	Background factors X **Item 9:** The synthesised voices were sufficiently clear to make the speech intelligible. **Item 11_2:** Please give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here?
Hypothesis 3		**Theoretical Background**	**Questionnaire Items**
The synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness.		Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	**Item 11_2:** Please give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here?

		Item 11_3: Please give your opinion on the attractiveness of the voices.
Hypothesis 4	**Theoretical Background**	**Questionnaire Items**
A majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented.	Described in Chapter 3.3.2.2: 'Criteria for Evaluating the Irish TTS synthesis in the Present CALL Platforms'	**Item 11_2:** Please give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here?
Hypothesis 5	**Theoretical Background**	**Questionnaire Items**
There is a statistically significant relationship between the measured reactions of the respondents to the non-TTS elements of the platform and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform.	See discussion in Chapter 2.5.6, in particular "...game playing is a holistic experience" "...one's view of particular aspects of the game, for example graphics, plot (or TTS voices) is influenced by one's disposition towards the whole of the game..."	All post-game questionnaire items
Hypothesis 6	**Theoretical Background**	Questionnaire Items
There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general & their attitude towards the Irish synthetic voices being tested.	See discussion in Chapter 3.3.2.1.1 and "...evidence to suggest that one's prior experience of synthetic voices may have an influence on one's attitude towards a synthetic voice in a particular context (Cryer & Home, 2009)"	Background factors X **Item 9:** The synthesised voices were sufficiently clear to make the speech intelligible. **Item 11_2:** Please give your opinion on the quality of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here? **Item 11_3:** Please give your opinion on the attractiveness of the voices.

The questionnaire results are presented in Section 5.9 and the implications of these results for the hypotheses of this study are discussed in Section 5.9.4 where the findings relevant to each of the evaluation criteria and the questionnaire items are dealt with in turn.

5.5 Procedure

As was the case with the teacher evaluations, initial contact was made with teachers who attended an Irish language teaching conference in NUIM, Maynooth ('An Tionól Teagaisc' in 2011). I explained to teachers that day that I would like to have senior cycle second level learners of Irish participate in the research and I showed the teachers some excerpts from the games. There was great enthusiasm for the study and teachers gave me their contact details. Invitations to take part in the present study were extended initially by email to teachers I had met at the conference. After some contact regarding practical issues, such as availability of computer rooms, internet connectivity and functioning sound cards in the machines, I followed up with teachers who were in schools which had the required facilities and where they could allocate sufficient time to a group to participate in the study. It was considered desirable that various backgrounds would be represented in this mix - i.e. schools from each of the four provinces in a mixture or urban and rural settings.

The system was piloted initially in one school with 30 pupils and that led to some restructuring of questions based on feedback concerning possible ambiguities. The pilot test was also useful from the point of view of working out the timing and of establishing how best to apportion the available time to the games and to the questionnaires. Teachers in this school also participated in the initial pilot testing. The test results from the school have not been included in the overall results since the questions and the form of the questionnaires were not identical to those given to the main groups.

Students took part during their school day and used the school's computing facilities. I provided headphones for each student so as to ensure that all had relatively high quality access to the aural material. No time limit was given to the students but most teachers agreed to set aside three class periods (of 40 minutes each). In some cases, these three class periods ran as a single session and in others, students were given a double period (80 minutes) first, and a single class period at another stage during the same day to take part in the evaluations. This gave ample time to fill in the pre-game questionnaire, to play each of the games and to fill in a post-game questionnaire after each of the three games.

I was present at each of the evaluations and began the session by introducing the *ABAIR* project and explaining what TTS synthesis entailed. I told students that this was a university-based research project. I explained that the main aim of the *ABAIR* research project was to develop synthetic voices for Irish and that I was particularly interested in whether these voices could be used in the context of language learning. I explained that the games had been developed in a university context and that I was interested in their opinions of various aspects of the games in terms of language learning in the Irish context. I invited questions from students and addressed any questions/comments they had.

The evaluations took place in a relaxed atmosphere and students were free to ask questions about any aspect of the study as they progressed through the games. In order to offer increased anonymity in the present study, usernames ('user101'; 'user102'; etc.) were randomly assigned to each of the respondents obviating the need to use their actual names. They were free to opt out of the research if they so wished and were told that they could continue with their own school work quietly if they did not wish to participate. Interestingly, none of the students opted out of the study. Most students took part in the evaluations of all three platforms, with the exception of a small number who had other commitments during the school day (such as sporting events / medical appointments) and who had to leave at a specified time. I was present for the entire duration of the evaluations. Apart from spending time with individuals answering their questions, I

ensured that I was either moving around the room or in a particular place in the room (depending on layout) where no student felt that I was looking over his/her shoulder at the responses. Students were given a break between each of the games.

The questionnaires administered to the pupils are included in Appendix C (pre-game questionnaires) and Appendix E (post-game questionnaires, including results). The evaluations were all delivered online. The online delivery presented a number of technical difficulties, such as inability to access YouTube (where *Fáilte go TCD* was being hosted) in some schools and other general maintenance problems associated with computers in schools. These problems were overcome but they are a reminder of the difficulties that can exist with the whole concept of internet-based CALL throughout the country at present.

5.6 Analysis

This section of the present study deals with subjective data elicited from learners by way of Likert scale-based questionnaire items.

Chapter 3.3.4 gave an overview of the statistical analysis used, beginning with a background review on Likert scales. It went on to outline methods of statistically analysing data from Likert scales and argued that use of non-parametric procedures was the most appropriate in the context of this study.

For analysis purposes, data gleaned from the pre-game questionnaire (described in Section 5.3.2.1) were regarded as independent variables and pupils' opinions, gleaned from post-game questionnaires (described in Section 5.3.2.2), were regarded as dependent variables.

A number of key points of information, in relation to the backgrounds of the respondents, were established from data gathered from the pre-game questionnaire administered to students. Based on those data, the following independent variables are examined and findings will be reported in Section 5.9:

1. Gender Differences
2. Between-School Differences
3. Frequency of Playing Computer Games
4. Level of Understanding of Irish
5. Attitude towards Synthetic Voices

The responses to each of the post-game questionnaire items are regarded as the dependent variables. The scores given by respondents to an item on the post-game questionnaire, such as their opinion of the graphics, might or might not have a statistically significant relationship with an independent variable such as gender, school type, etc. The purpose of the analysis is to test the null hypothesis - that variation in the dependent variables is not significantly related to variation in the independent variables. Put another way, it seeks to answer whether 'gender, school type, level of Irish, frequency of playing computer games, attitude towards synthetic voices' have a statistically significant impact on the respondents' rating of TTS synthetic voice, usability of game, impact of graphics, etc?

Chapter 3.3.4.4.1 described the Kruskal-Wallis Test. The Kruskal-Wallis test was identified as being the most suitable at testing for statistical relationships between various background characteristics (the independent variables), such as gender or the school type which one attends, on the one hand, and pupils' responses to various items which indicated their attitude towards the platforms being, evaluated on the other.

The Spearman's Rank Correlation Coefficient (Spearman's rho) is described in Chapter 3.3.4.4.2. A Spearman's rho analysis technique was identified as being the most suitable test to examine the relationship between pupils' responses to different aspects of the platforms, including their TTS and non-TTS features. These data test the extent to which playing the games is a holistic experience and whether one's response to one aspect of the game is influenced by one's reactions to other aspects of the games. As has been repeatedly pointed out, one's view of the intelligibility, quality and attractiveness of the synthetic speech could be significantly related to one's overall experience of playing the games.

The data presented in Chapter 5.9.3.2 are in their raw, or descriptive format, i.e. the raw data are shown in straight numerical and percentage terms. In Chapter 5.9.3.3 inferential statistics are used in order to test for statistically significant relationships between the background categories mentioned above, or independent variables, and pupils' responses to the various post-game questionnaire items, or dependent variables. The statistical procedure used to test these relationships is the Kruskal-Wallis test. Spearman's rho is also calculated in order to establish the correlation between responses to the post-game questionnaire items. This tests the contention that game playing is a holistic experience and that one's perception of individual facets of the experience is governed by one's overall impression. All calculations were made using the IBM SPSS Statistics package, Version 19.

5.7 Reliability and Validity

Reliability and Validity were taken into consideration at the design stage of this study. Some of the main features are summarised in this section.

Chapter 3.3.4.2 discusses lack of anonymity as being a potential risk to the validity of the responses to Likert scale-based questionnaire items (McLeod, 2008) insofar as respondents might wish to give the 'polite' answer or, tend to answer towards the

midpoint of the scale (Worcester & Burns, 1975). One possible safeguard to counter these risks is to offer anonymity to the respondents. In the present part of the study respondents answered using online questionnaires. As described in Section 5.5 above, students were offered anonymity in that a username was ascribed to each participant, as opposed to using their own name. Additionally, although I was present when the evaluations were being carried out, I ensured that I was either at the top of the classroom or moving around the room so that no student felt that he/she was being watched while answering the questionnaire items.

The validity of the Likert scale ratings given to questionnaire items is considered to be high because of the very short timespan between pupils experiencing the game-like CALL platforms and their completing the post-game questionnaires. If there was any delay in administering the post-game questionnaire, it could be argued that a certain memory loss would take place which would, in turn, reduce the validity of the subjective responses. Validity is also considered high in this study since participants are all of a similar age, which is the actual age of the target population for these platforms.

Generalisability is enhanced by the fact that there is a large number of students taking part, who are of the same age group, representing a rather large geographical and sociolinguistic spread.

5.8 Ethics

A request for ethical approval for the present research project was submitted to the Research Ethics Committee of the School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin for consideration.

This process involved submitting:

- The Research Ethics Application Form of the School of Linguistic, Speech & Communication Sciences (in English only);
- a letter to school principals seeking access to the second level schools selected for this study (both in English and in Irish);
- an information leaflet with information for participants (both in English and in Irish);
- consent forms for pupils (both in English and in Irish);
- letters to parents with a description of the research being conducted (both in English and in Irish);
- parental consent forms for students under 16 years (both in English and in Irish);
- copies of student pre-game and post-game questionnaires (both in English and in Irish);
- a description of the three platforms being evaluated (in English only);
- a timeline for the research project (in English only);
- a timeline for when students/teachers would be contacted (in English only);
- a description of the procedure envisaged (i.e. where the research would take place; time allowed; whether the researcher would be present, what would happen in the event of students wishing to opt out of the research, etc.) (in English only);

After some minor revisions (mainly involving some simplification of the wording in the student information leaflets), the project was approved in full.

5.9 Results

5.9.1 Overview.

'Results' refers to the conclusions arrived at, based on the analysis of the data, as to whether the hypotheses laid out first in Chapter 3.4 can be supported.

Section 5.9.2 details the results of the pre-game questionnaire which elicited background information on the respondents. This information is necessary in order to test the relationship between these variables and the opinions expressed on the various aspects of the platforms, including the synthetic speech.

Reactions to each of the three platforms were measured by 'post-game' questionnaire items. For the purpose of analysis, results for each of the platforms are taken separately. Also separated, are the reactions of pupils to the synthetic speech (TTS) and to all other (non-TTS) aspects of the platforms. Results are given here in summary form and full details are laid out in Appendices E1, E2 and E3. Results are initially set out as descriptive statistics which give the numbers and percentages of participants who responded to the 'post-game' questionnaire items. This part of the analysis is presented in Section 5.9.3.2.

Section 5.9.3.3 brings the statistical analysis to an inferential level where background variables amongst the respondents are correlated with their responses to the post-game questionnaire items. This is done by using the Kruskal-Wallis analysis in order to test for influences of background factors on response patterns and results are presented separately for each of the three platforms.

This section also applies the Spearman's rho correlation technique in order to test the hypothesis that there will be a significant relationship between response patterns to each of the items in the post-test questionnaire. A strong correlation between the items would

suggest that pupils' engagement with each platform is a holistic experience and that taking a positive view of one aspect of the platform will lead to one taking a positive attitude to many other aspects of the platform. This test also allows us to draw inferences on the extent to which evaluators' ratings of the TTS aspects of the platform are influenced by their reaction to the non-TTS aspects, such as graphics, plot, etc.

Figure 5.2 gives a visual overview of the types of analyses (descriptive and inferential statistical) which are reported on in this section.

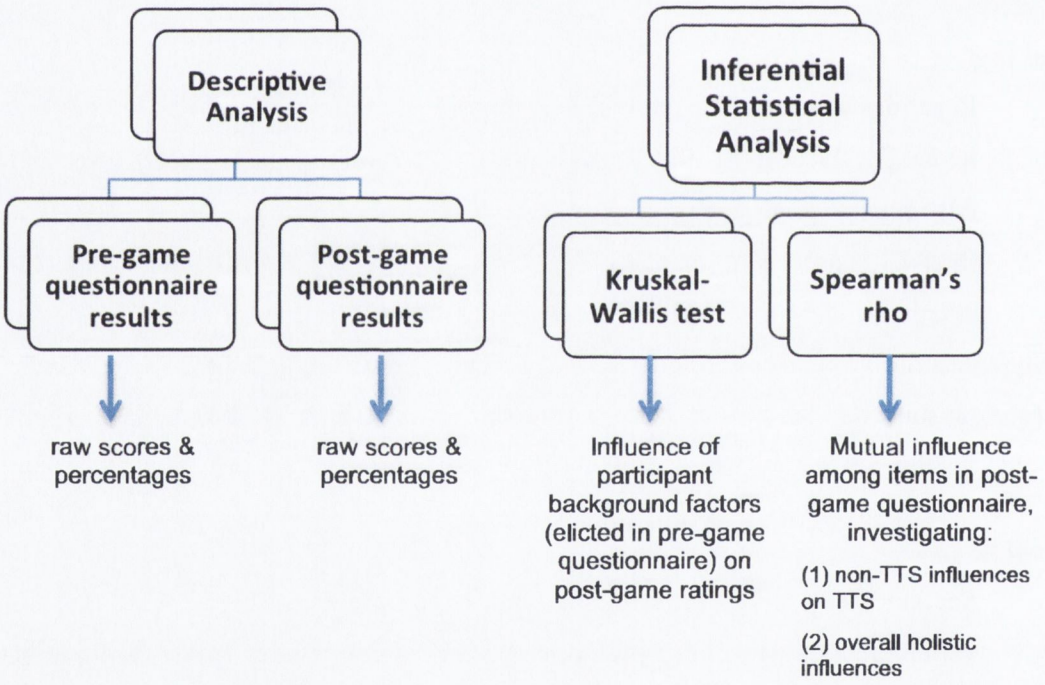


Figure 5.2. Overview of descriptive and inferential statistical analyses.

In Section 5.9.4 results from the previous sections are interpreted in the context of the hypotheses stated in Chapter 3.4 and there is a discussion on the extent to which the hypotheses can be supported in the light of these results. A summary and brief discussion are presented in Section 5.9.5.

5.9.2 Learner pre-game questionnaire.

Table 5.6 below illustrates the background factors pertaining to pupils, as elicited from the 'pre-game' questionnaires for each of the three platforms. As can be seen from Item 1, the numbers of participants varied slightly from test to test for the reasons that were mentioned earlier – e.g. medical appointments. It can be seen from Item 2 that there were considerably larger numbers of females than males.

The results for the second platform, *Fáilte go TCD,* are described here. This platform was chosen for the commentary simply because it was the platform evaluated by the largest number of participants. Although the commentary is based on the second platform, the results were extremely close for all three platforms and so individual comments capture the relevant trends across the different platform evaluations.

Gender (Item 2)

Item 2 shows that there were considerably more girls than boys in the overall cohort (181 female: 71 male).

School Year (Item 3)

The majority of pupils who took part in the evaluations were in their 5th year in second level schools, putting them in the 17 year old age bracket. Only 22 of those who took part in the *Digichaint* and *Taidhgín* evaluations were in 4th year and a further 3 were in 6th year. Given the very small numbers not in 5th year, this item was not included in the statistical analysis.

School-Type (Item 4)

In terms of school type, 144 pupils came from English-medium schools with a mixture of urban and rural backgrounds, 23 came from urban Gaelscoileanna and 85 came from rural Gaeltacht schools. This is not a representative sample of the student population of schools in Ireland. Gaeltacht schools are deliberately over-represented as it was felt that students from these schools were in a somewhat better position to judge the clarity of the synthetic voices. One's ability to rate synthetic voices is strongly dependent on the listening comprehension ability of the participant (Kang et al., 2008). If one's listening comprehension ability is below the level of the material being presented then one's judgment of the clarity of the voice is suspect. A more detailed discussion on school selection is included in Section 5.2 above.

Frequency of Playing Computer Games (Item 5)

In Item 5 of the questionnaire, participants were asked to indicate how often they played computer games generally. The modal response to this question was 'rarely' (55%) and a further 19% reported that they never play computer games. Only 8% reported that they played computer games on a daily basis while the remaining 18% reported that they played on a weekly basis.

Level of Understanding of Irish (Item 6)

Item 6 of the pre-game questionnaire asked participants to rate themselves on their level of Irish language comprehension. 78% reported that they could understand 'almost all conversations when spoken at a normal speech rate' or 'almost all conversations with clear pronunciations'. 19% said they could understand 'pieces of conversations' while the remainder rated themselves as understanding 'just a few words or a few simple sentences'. 94% of participants had taken the higher level Junior Certificate paper in Irish and 91% intended to take Irish at higher level in the Leaving Certificate examination.

Attitude towards Synthetic Voices (Item 10)

Participants had a somewhat neutral attitude towards synthetic speech in general. 51% reported that they didn't mind the use of synthetic voices as long as they were intelligible.

A further 22% had a more negative attitude indicating either that they would prefer the use of human voices or that they hated synthetic voices. A similar number had a positive attitude towards synthetic voices, as shown by the remaining 27%, who said that they were sometimes as appropriate or more appropriate than a human voice.

Level of Junior Certificate and Leaving Certificate examinations (Items 7 and 8)

Two further items were included in the questionnaire for which results are presented in Table 5.6. These were the level of the Junior Certificate examination (a State examination taken at approximately 16 years of age) which participants had taken (Item 7), and the level which participants anticipated taking in their Leaving Certificate Irish (pre-University) examination (Item 8). The overwhelming majority had taken, or intended to take, the higher level in both examinations, although the numbers for the Leaving Certificate were slightly lower than those for the Junior Certificate. Nonetheless, given the very high percentages taking the higher level, these items are not included in the statistical analysis.

Ranking of 'Top 3' school subjects (Item 9)

Participants were asked to rank their 'Top 3' school subjects in order of preference. This item was included in order to pick up on positive attitudes of participants towards the learning of Irish: a high placing for Irish in their subject preference ranking would indicate a positive attitude. Students at this level generally study 7 or 8 subjects for Leaving Certificate. Findings on this evaluation item proved to be unsatisfactory in a number of ways. This was due to the way in which the answer was elicited in an open dialogue box. Although 45% ranked Irish among their top three favourite subjects, there were large numbers who filled in the box in erratic ways, e.g. by listing non-curricular items such as sports, by failing to list any subjects, etc. Therefore, while the positive finding of 45% was noted, one cannot make too many assumptions about overall attitudes. It was not possible to use this item in subsequent statistical evaluations.

Table 5.6

Pupil pre-game questionnaire results for all three platforms (Digichaint, Fáilte go TCD and Taidhgín)

			Digichaint		Fáilte go TCD		Taidhgín	
			Number	**Percent.**	**Number**	**Percent.**	**Number**	**Percent.**
1	Total Number of Participants		250	100%	252	100%	228	100%
2	Gender	Female	182	73%	181	72%	166	73%
		Male	68	27%	71	28%	62	27%
3	Year of Study	4th	22	9%	39	15%	22	10%
		5th	225	90%	213	85%	203	89%
		6th	3	1%	0	0%	3	1%
4	School type	Gaeltacht	86	34%	85	34%	80	35%
		Gaelscoil	25	10%	23	9%	22	10%
		English-medium	139	56%	144	57%	126	55%
5	Frequency with which students play computer games	Never	49	20%	48	19%	47	21%
		Seldom	133	53%	139	55%	127	56%
		Weekly	50	20%	45	18%	37	16%
		Daily	18	7%	20	8%	17	7%

			Digichaint (continued)		Fáilte go TCD (continued)		Taidhgín (continued)	
			Number	**Percent.**	**Number**	**Percent.**	**Number**	**Percent.**
6	Self-evaluation of comprehension competence	Only a few words spoken slowly	2	1%	3	1%	3	1%
		A few simple phrases spoken slowly	3	1%	6	2%	2	1%
		Parts of conversations	38	15%	47	19%	36	16%
		Most conversations when spoken clearly	114	46%	108	43%	100	44%
		Almost all conversations at natural conversation speed	93	37%	88	35%	87	38%
7	Junior Cert. Exam level	Higher	240	96%	238	94%	224	98%
		Ordinary	10	4%	14	6%	4	2%
8	Leaving Cert. Exam level	Higher	233	93%	230	91%	219	96%
		Ordinary	17	7%	22	9%	9	4%
9	Irish in 'Top 3' favourite subjects		110	44%	113	45%	101	44%
10	Attitude towards synthetic voices	I hate them	6	2%	13	5%	10	4%
		Tolerate them but prefer human voices	47	19%	42	17%	45	20%
		Neutral	144	58%	129	51%	111	49%

		Sometimes Suitable	48	19%	63	25%	60	26%
		Sometimes more suitable than human voices	5	2%	5	2%	2	1%

5.9.3 Learner post-game questionnaire.

5.9.3.1 Overview.

This section is divided into two main parts – the first, which is termed 'descriptive results', reports the numeric results obtained from the questionnaires in a purely descriptive manner as raw scores and percentages, and is presented in Section 5.9.3.2. The second part (see Section 5.9.3.3) deals with the inferential statistical analysis. This looks at the extent to which the participants' background factors (elicited in the pre-game questionnaires) influence the judgments made in the post-game questionnaires, using the Kruskal-Wallis test. It also uses Spearman's rho to examine the extent to which there is a broad interdependency between all questionnaire items in a way that would suggest that these platforms are being experienced in a holistic manner. This analysis is presented in Section 5.9.3.3.6. For ease of reference, 'inferential statistical analysis' will often be simply referred to as 'statistical analysis'.

This Results Section sets out to address the research questions, as set out in Chapter 3.4. The full detailed results of the post-game questionnaires are presented in Appendices E and H. Appendix E provides all detailed results relevant to the descriptive account: Appendix E1 refers specifically to *Digichaint,* Appendix E2 refers to *Fáilte go TCD* and Appendix E3 to *Taidhgín*. They are presented separately even though there is some overlap in the questions relating to each of the three platforms. Note, in particular, that *Digichaint* has its own specific questions which relate to factors such as plot and

manoeuverability and which do not apply directly to the other two platforms. Appendix H provides a parallel description of the results summaries presented in this chapter for the inferential statistical analyses (see Section 5.9.3.3), providing detailed item-by-item analysis of the significant statistical findings.

5.9.3.2 Post-game descriptive results (non-TTS and TTS aspects).

As is mentioned in the 'Research Design and Strategy' section above (see Section 5.3.2.2), a distinction is being made here between learners' opinions on the general features of each of the platforms being evaluated (referred to as "non-TTS" aspects of the platforms) and information sought on the intelligibility, quality and attractiveness of the TTS synthetic voices ("TTS" aspects of the platforms). Sections 5.9.3.2.1, 5.9.3.2.2, and 5.9.3.2.3 below refer to findings on the non-TTS features of the platforms, such as plot, graphics and manoeuverability. The results of questionnaire items specifically referring to the intelligibility, quality and attractiveness of the TTS synthesis are set out separately in Section 5.9.3.2.4. The results are presented in terms of the raw data and percentages. In both subsections, results are discussed for each of the three platforms in turn.

Before proceeding to the more detailed discussion of results it should be noted that there is a reasonable spread of responses on the Likert Scale to each of the questionnaire items. This suggests that there was an honesty in the responses and that learners did not opt merely for the 'polite' answer (a potential risk mentioned in Section 5.7).

5.9.3.2.1 Non-TTS descriptive results: Digichaint.

The results of the evaluations set out in this section represent learners' reaction to each of the criteria for effective CALL platforms, as set out in Table 3.2 above. Details of the responses to each of the questionnaire items are given in Appendix E1.

Overall reaction to the *Digichaint* game can be described as '*positive*' or '*very positive*'. Learners' attention was primarily directed at the game plot and at playing the game (54% selected '*agree*' or '*agree completely*' for Item 9: 'I was more focused on the plot of the game than I was on the language being used'). The game was regarded as having a good balance between playfulness and opportunities for language learning (75% '*agreed*' or '*agreed completely*' for Item 10). Most claimed that they learned some new words or phrases (66.4%: *note that this included Gaeltacht and Gaelscoil pupils who would be likely to have a higher standard of Irish than that in the game) and that the language difficulty level was appropriate (70%). The game plot was regarded as clear (77% '*agreed*' or '*agreed completely*' with Item 6), realistic (68%: Item 7) and pleasant to play (69% gave a positive response for Item 5). Learners engaged with the game quite intensively and remained concentrated for its duration (66% selected points 4 and 5 on the Likert scale stating that the game 'held their attention'). The visual aspect of the game was considered acceptable by the majority (60%: Item 1) but it is clear that they are not considered to be of the same standard as commercial games.

5.9.3.2.2 Non-TTS descriptive results: Fáilte go TCD.

Details of the responses to each of the questionnaire items for the *Fáilte go TCD* platform are given in Appendix E2. 61% of respondents found the graphics '*attractive*' or '*very attractive*' while 23% thought the opposite (Item 1). Only 54% thought that the speech was '*well aligned*' or '*very well aligned*' to the movements (Item 2). There was a similar low acceptance of the contribution that the graphics made to the credibility and clarity of the speech (51% '*good*' contribution or '*very good*' contribution (Item3)).

While there was a low acceptance of the graphics there was, nevertheless, a strong agreement that this type of environment would be '*helpful*' or '*very helpful*' for the development of language comprehension (70%: Item 4). 75% of respondents said that they would '*enjoy*' or '*very much enjoy*' using this type of facility should it be readily available. 63% considered it to be '*motivating*' or '*very motivating*' (Item 6) and 77% '*agreed*' or '*strongly agreed*' that it would make the learning of Irish more attractive.

70% thought that the language was at an appropriate level of difficulty and, of those who did not, only 31% thought it to be too difficult, i.e. 13% of the overall group (Items 8.1 and 8.2).

5.9.3.2.3 Non-TTS descriptive results: Taidhgín.

While *Taidhgín* had the least complex graphics it had the highest attraction for respondents. 87% considered the graphics to be '*attractive*' or '*very attractive*', while only 6% found them '*unattractive*' or '*very unattractive*' (Item 1). 90% thought that the monkey added to the playfulness of the platform (Item 2). 68% considered that the movements of the monkey were '*well aligned*' or '*very well aligned*' to the speech with only 12% of the opposite opinion. Overall one can consider the development of the graphics on the platform to have been successful.

86% reported that they would find this type of program '*useful*' or '*very useful*' in helping them practise conversational Irish (Item 4) while only 5% considered it to be of 'little or no use'. 79% claimed that they would use such a platform should it be readily available. 82% thought it would be motivating (Item 6) and 88% thought it would make the learning of Irish more attractive.

Overall one can say that the platform got an extremely high approval rating in terms of its graphics and its general attraction for respondents. Appendix E3 illustrates in detail the items asked and the breakdown of the responses to the *Taidhgín* platform.

5.9.3.2.4 TTS descriptive results for the three platforms.

This section focuses on an evaluation of the TTS synthetic voices which were integrated into each of the three platforms. Focus is being directed on three aspects of the synthetic voice which can be compared across all three platforms – intelligibility, quality and attractiveness. The response data are included in Appendices E1, E2 and E3 for the

individual platforms but the TTS responses are being brought together here to allow for comparison in Table 5.7 which sets out the percentage of respondents who expressed a '*positive*' or '*very positive*' disposition towards the voices, i.e. they marked them at points 4 or 5 on the Likert scale.

Table 5.7

Responses to the TTS aspects of the three platforms: percentage of pupils who were 'Positive' or 'Very Positive' towards the Intelligibility, Quality and Attractiveness of the Synthetic Voices

	Digichaint	***Fáilte go TCD***	***Taidhgín***
The synthesised voices were sufficiently clear to make the speech **intelligible**.	(Item 12) 56%	(Item 9) 45%	(Item 9) 73%
Give your opinion on the **quality** of the synthesised voices: do you think the voices are adequate for the type of game presented here?	(Item 17_2) 62%	(Item 11_2) 51%	(Item 11_2) 73%
Give your opinion on the **attractiveness** of the voices:	(Item 17_3) 42%	(Item 11_3) 38%	(Item 11_3) 57%

Initial inspection of results shows a relatively low rate of positive or very positive responses to the intelligibility, quality and attractiveness of the speech. It may also be noted that the judgments based on the *Taidhgín* platform were very much higher than those of the other two. It is noted that the *Taidhgín* platform uses only the male HTS voice in its 'original' form (without any pitch and speed manipulations). This is because only one character is presented in this platform. (Pitch and speed manipulations to the 'original' voice were used as the solution to differentiate between voices when more than one character was speaking in the other two platforms, see Chapter 3.2.1.2. In *Taidhgín*, students were also given the option of hearing their own input spoken aloud by either the female Ulster-dialect (unit-selection) voice or the male Connaught-dialect (unit-selection) voice – both of which were in their 'original' synthetic forms, without any crude pitch

and speed alterations as was necessary to include in both *Fáilte go TCD* and *Digichaint* because of the extended cast of characters in these platforms.

Item 14 of the *Digichaint* questionnaire asked respondents to react to the proposition that it was no more difficult to understand the computer-generated voice than natural voices. 49% '*agreed*' or '*strongly agreed*' with the proposition while 34% expressed the opposite point of view. On the question of synthetic voices giving a better atmosphere to computer games than natural voices can do 37% '*agreed*' or '*strongly agreed*' with this while 40% '*disagreed*' or '*strongly disagreed*'. This is the only issue on which a negative response outweighs the positive one even if the margin is narrow.

5.9.3.2.5 Summary of descriptive analysis.

An inspection of the descriptive data, as laid out in Appendices E1, E2 and E3, shows an overall positivity towards each of the three platforms. The numbers expressing a '*negative*' or '*very negative*' attitude as a response to any of the questionnaire items were very small. It was against this positive attitude towards the platforms that evaluations of the synthetic speech were made. The 'attractiveness' of the speech got the lowest rating suggesting that natural speech is still very much preferred to synthetic speech. The *Fáilte go TCD* speech got the lowest overall rating for its intelligibility, quality and attractiveness. The *Taidhgín* platform got the highest rating. Since the voices were from the same *ABAIR* system, it seems reasonable to suggest that when there is a level of enjoyable interactivity between player and platform then there will be a higher estimation of the quality of the synthetic speech. *Fáilte go TCD* is a relatively inactive platform compared to the other two. This is likely to have influenced ratings and may explain the relatively lower ratings.

5.9.3.3 Post-game inferential statistical results.

5.9.3.3.1 Overview.

This section presents the results of an indepth statistical analysis of the raw data presented in Section 5.9.3.2. Statistical analysis techniques used in this chapter were introduced in Chapter 3.3.4 of this thesis. In the first part of the analysis, the Kruskal-Wallis test is used in order to further analyse the influence of participant background factors (elicited in pre-game questionnaire) on their responses to the post-game questionnaire items. Analysis for each platform was conducted separately and is presented in terms of the relationship between background factors (i.e: gender; school type; frequency playing computer games; level of understanding of Irish and predisposition towards synthetic voices (independent variables)) and pupils' reaction to post-game questionnaire items (dependent variables).

The analysis is carried out in order to test the null hypothesis that there is no relationship between the background factors listed and pupils' responses to the post-game questionnaire items. In this case, I have, a priori, proposed two hypotheses which can be statistically tested. These hypotheses were introduced in Chapter 3.4 and are:

- Hypothesis 2: One's standard of Irish comprehension is significantly related to one's opinions on the intelligibility, quality and attractiveness of the synthetic speech – ("the synthesised speech is readily intelligible for respondents with a high Irish language comprehension ability")

- Hypothesis 6: There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general and their attitude towards the Irish synthetic voices being tested in the three newly developed CALL platforms.

Hypotheses 2 and 6 are being tested by the Kruskal-Wallis test, a non-parametric equivalent of a standard one-way ANOVA, which is used when data are from a suspected non-normal population. This omnibus procedure tests for some differences between three or more groups (Elliott & Hynan, 2011).

In the second part of the statistical analysis, the internal correlations between the post-game questionnaire items are analysed in order to test Hypothesis 5:

- Hypothesis 5: All post-game items are interconnected – ("there is a statistically significant relationship between the measured reactions of the respondents to the non-linguistic elements of the platform (such as plot, ease of manoeuvring characters, legitimacy of narrative and standard of Irish language used) and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform").

It was an expectation that rating post-game questionnaire items might be done in a rather holistic manner and, so, a positive response to one aspect of the platform might positively influence responses to another. Of particular interest was to investigate whether ratings given to the synthetic speech (TTS) were to some extent dependent on ratings given to other aspects of the platform (non-TTS aspects). This statistical analysis was done by way of Spearman's Rank Correlation Coefficient (Spearman's rho). This is a technique used with ordinal or ranked data and it is implemented in order to determine the strength of the relationship between two measurements.

Returning to the first part of the analysis, the Kruskal-Wallis test, where respondents' background factors were correlated against all post-game questionnaire items, it should be noted that three of the background factors were excluded, following preliminary inspection of the data, as either irrelevant or unworkable. These items were 'year in school', 'honours/ordinary level at Junior/Leaving Certificate examinations' and 'Top 3 favourite subjects'. These items have been discussed in Section 5.9.2.

The inclusion of all other participant factors in the Kruskal-Wallis test means that a number of correlations are being calculated which can allow insights beyond what is being directly sought as support for the hypotheses. Such calculations are referred to as post hoc analyses (Ruxton & Beauchamp, 2008; Thomas, 1997) and involve looking in the data for patterns that were not specified a priori. There is a certain danger in making claims based on any such analyses as their generalisability to a wider population may be less robust insofar as they could be due to chance characteristics of the particular group being tested. It was felt important, nonetheless, to include discussion of such broader potential influences and the results for these items are being presented in a more tentative way, as they may be informative rather than carrying the normal statistical reliability levels.

The results for the Kruskal-Wallis test concerning each of the three CALL platforms are presented in Section 5.9.3.3.2: *Digichaint* statistical analysis; Section 5.9.3.3.3: *Fáilte go TCD* statistical analysis; and Section 5.9.3.3.4: *Taidhgín* statistical analysis, and implications for the hypotheses are discussed in Section 5.9.4. A general commentary on each of the significant results is given under five specific headings corresponding to the background factors examined and more detailed item-by-item analyses of those significant results are included separately in Appendix H. The five background factors included in the analysis are:

1. Gender
2. Between-School Differences
3. Frequency of Playing Computer Games
4. Level of Understanding of Irish
5. Attitude towards Synthetic Voices

For items which showed significant statistical relationships a further analysis was carried out, also by way of the Kruskal-Wallis test, in order to pinpoint where precisely the significant relationships occurred. For example, was there a significant difference between each of the three types of schools being examined, i.e. Gaeltacht schools,

Gaelscoileanna and English-medium schools or was it the case that there was no difference between the first two and the third was significantly different from both?

The direction of the differences established by the Kruskal-Wallis test can be noted from the mean rank score. These mean rank scores are reported in this chapter (and in the accompanying appendices) only for statistically significant differences in the initial analysis. Commentary is given on where the points of significance exist (i.e. whether there is a significant difference between each of the categories being analysed or whether the difference affects only some of the categories being tested). Full numeric details on the analysis are available in tabular form in Appendix F1 (*Digichaint*), Appendix F2 (*Fáilte go TCD*), Appendix F3 (*Taidhgín*). Appendix H provides a parallel extended discursive presentation of the accounts in Sections 5.9.3.3.2 – 5.9.3.3.4 for each of the three platforms separately under the five specific headings (1) 'Influence of *Gender*' to (5) 'Influence of *Attitude Towards Synthetic Voices*'. These appendices are also referred to in the discussion below.

Tables 5.9, 5.11 and 5.13 below set out the results of the Kruskal-Wallis analysis for *Digichaint, Fáilte go TCD* and *Taidhgín*, respectively. Significant statistical relations are in red font and are marked with an asterisk (*).

5.9.3.3.2 Digichaint statistical analysis.

This section discusses the results of the Kruskal-Wallis test which is used here to test the significance of the relationship between respondents' background factors (established in the pre-game questionnaire, Section 5.9.2/Table 5.6) and post-game questionnaire items for the *Digichaint* platform. Summaries of the statistically significant results follow, more detailed item-by-item analysis is included in Appendix H and the implications of these results for the hypotheses are collated in Section 5.9.4.

The precise post-game questionnaire items are included in Appendices C2 (in Irish) and E1 (in English) but are reproduced in Table 5.8 below for convenience.

Table 5.8

Summary of Digichaint post-game questionnaire items – those referring to TTS aspects are shaded.

Item 1	Rate the graphics used in this game on a scale of 1 – 5.
Item 2	Did you feel the game was technically very difficult to play – very easy to play?
Item 3	I learned some new phrases/words/grammar points as I played the game.
Item 4	The dictionary is helpful as an aid to learning Irish in this game.
Item 5	Rate your overall enjoyment of this particular game as a language learning experience.
Item 6	The plot of the game was clear to me from the outset.
Item 7	Was the plot credible for this type of game?
Item 8	The game held my attention.
Item 9	I was more focused on the plot of the game than I was on the language being used.
Item 10	There is a good balance between enjoyment and language learning in this game.
Item 11	The overall standard of the Irish used in this game is at about the right level for me.
Item 12	The synthesised voices were sufficiently clear to make the speech **intelligible**.
Item 13	Did you experience particular difficulties with the dialects that are used in Digichaint?
Item 14	I found it no more difficult to understand the computer-generated voice than I would if natural voices were used.
Item 15	In general, I think that computer-generated voices give a better atmosphere to computer games than natural voices do.
Item 16	I found the computer-generated voice suitable for this computer game
Item 17_1	Give your opinion on the usefulness of the concept of producing an interactive language learning game in order to practise Irish.

Item 17_2	Give your opinion on the **quality** of the synthesised voices: to what extent do you think the voices are adequate for the type of game presented here?
Item 17_3	Give your opinion on the **attractiveness** of the voices.

Table 5.9 below sets out the results of the Kruskal-Wallis analysis for *Digichaint*. The influence of the five participant background factors are then summarised one by one.

Table 5.9

Results of the Kruskal-Wallis analysis of the Digichaint Post-Game Questionnaire Items and Participant Background Factors. Statistically significant differences at $p<0.05$ are denoted in red font with an asterisk. Questionnaire items referring to TTS aspects are shaded.

Item	Gender	School Type	Frequency playing computer games	Level of understanding of Irish	Attitude towards synthetic voices
1	**H(1)=6.120, p=0.013***	**H(2)=23.133, p=0.000***	H(3)=4.210, p=0.240	**H(4)=17.629, p=0.001***	H(4)=4.420, p=0.352
2	H(1)=0.041, p=0.839	H(2)=2.327, p=0.312	H(3)=0.737, p=0.737	H(4)=6.175, p=0.186	H(4)=5.747, p=0.219
3	H(1)=3.713, p=0.054	**H(2)=27.139, p=0.000***	H(3)=6.354, p=0.096	**H(4)=11.978, p=0.018***	H(4)=8.556, p=0.073
4	H(1)=3.780, p=0.052	**H(2)=26.624, p=0.000***	**H(3)=8.656, p=0.034***	**H(4)=12.280, p=0.015***	H(4)=3.015, p=0.555
5	H(1)=3.360, p=0.067	**H(2)=29.324, p=0.000***	**H(3)=7.989, p=0.046***	**H(4)=12.972, p=0.011***	**H(4)=14.118, p=0.007***

Item	Gender (*continued*)	School Type (*continued*)	Frequency playing computer games (*continued*)	Level of understanding of Irish (*continued*)	Attitude towards synthetic voices (*continued*)
6	H(1)=0.291, p=0.590	H(2)=2.397, p=0.302	H(3)=4.140, p=0.247	H(4)=6.417, p=0.170	H(4)=0.573, p=0.966
7	H(1)=0.877, p=0.349	**H(2)=21.519, p=0.000***	H(3)=4.704, p=0.195	H(4)=8.612, p=0.072	H(4)=5.727, p=0.220
8	H(1)=1.639, p=0.200	**H(2)=19.302, p=0.000***	**H(3)=15.670, p=0.001***	H(4)=8.025, p=0.091	H(4)=6.590, p=0.159
9	H(1)=0.549, p=0.459	H(2)=2.457, p=0.293	H(3)=2.189, p=0.534	H(4)=5.366, p=0.252	**H(4)=11.963, p=0.018***
10	H(1)=0.094, p=0.759	**H(2)=6.792, p=0.033***	**H(3)=9.033, p=0.029***	**H(4)=9.800, p=0.044***	**H(4)=9.916, p=0.042***
11	H(1)=0.376, p=0.540	**H(2)=6.724, p=0.035***	H(3)=1.185, p=0.757	H(4)=8.304, p=0.081	H(4)=5.073, p=0.280
12	H(1)=0.070, p=0.792	H(2)=1.621, p=0.445	**H(3)=11.365, p=0.010***	H(4)=1.064, p=0.900	**H(4)=12.952, p=0.012***
13	H(1)=0.207, p=0.649	**H(2)=6.114, p=0.047***	H(3)=2.332, p=0.506	**H(4)=11.969, p=0.018***	H(4)=6.934, p=0.139
14	H(1)=1.027, p=0.311	**H(2)=6.739, p=0.034***	H(3)=2.660, p=0.447	H(4)=3.103, p=0.541	H(4)=7.342, p=0.119
15	H(1)=0.178, p=0.673	H(2)=0.939, p=0.625	H(3)=1.959, p=0.581	H(4)=6.166, p=0.187	**H(4)=11.785, p=0.019***

Item	Gender (*continued*)	School Type (*continued*)	Frequency playing computer games (*continued*)	Level of understanding of Irish (*continued*)	Attitude towards synthetic voices (*continued*)
16	H(1)=2.848, p=0.091	H(2)=5.001, p=0.082	H(3)=6.831, p=0.077	H(4)=6.237, p=0.182	H(4)=8.504, p=0.075
17_1	H(1)=,0.355, p=0.551	**H(2)=8.379, p=0.015***	H(3)=0.243, p=0.970	H(4)=2.500, p=0.645	**H(4)=16.076, p=0.003***
17_2	H(1)=1.281, p=0.258	H(2)=3.551, p=0.169	H(3)=7.562, p=0.056	H(4)=9.165, p=0.057	**H(4)=10.693, p=0.030***
17_3	H(1)=0.312, p=0.577	H(2)=0.714, p=0.700	H(3)=4.420, p=0.220	H(4)=1.991, p=0.737	H(4)=2.455, p=0.653

Influence of Gender

Gender showed very little impact on the post-game evaluation items. It was significant only in relation to respondents' opinions of the graphics where girls showed a significantly higher evaluation score than did boys.

Influence of *Between-School Differences*

School type had an influence on eleven of the nineteen variables. Gaelscoileanna and Gaeltacht schools formed a coherent group with respect to most items, with only four cases where statistically significant differences were found between them. This group seemed generally less impressed by the *Digichaint* experience than their counterparts in English-medium schools. This trend held for their opinions on the graphics, their belief that they learned new words or phrases from the game, their view on the usefulness of the

dictionary which was included as a hypertext link, their overall enjoyment of the experience, the credibility of the plot, the balance between enjoyment and language learning, the suitability of the language level, and the ease with which they could come to terms with the different dialects. Respondents from English-medium schools were more likely to agree that the synthetic speech was as clear to them as natural voices would be. They were also significantly more favourably disposed towards the concept of using interactive digital games for language learning. One may conclude that *Digichaint* was most attractive to those who found it a challenge and felt that their language learning experience could be enhanced by a digital game such as *Digichaint*. On the other hand, those with a more advanced knowledge of Irish found the experience less challenging and less rewarding. This is encouraging for the *Digichaint* platform since it shows that it was most attractive for the target group for which it was originally intended.

Influence of *Frequency of Playing Computer Games*

Playing computer games did not seem to occupy a great deal of the time of the respondents in this study. Only a minority reported that they played computer games regularly. Some familiarity with computer gaming seems to be an important factor in enhancing pupils' overall enjoyment of *Digichaint*. This may be because of the importance of familiarity with manipulating characters on a computer. Those who 'never' play games reported the lowest level of enjoyment from the game. One cannot safely speculate on the line of causality between 'never' playing computer games and lack of appreciation of *Digichaint*. It may be that some teenagers are just not interested in digital games of any kind and would not find them interesting as a means to learn any topic, including languages. Those who 'seldom' played games gave their enjoyment of the game the highest mean rank score. It is possible that those very familiar with gaming did not find the game challenging enough for them.

This same trend holds in the case of the game's ability to hold the players' attention. Those who 'never' play computer games had the lowest ranking for the game's ability to hold their attention. They also had low ratings for the balance between enjoyment and

language learning potential of the game and for the clarity of the synthetic voices. Overall, those with some experience in computer gaming enjoyed *Digichaint* the most while regular gamers were somewhat less impressed by it.

Influence of *Level of Understanding of Irish*

The relationship between pupils' level of understanding of Irish and the post-game items shows a similar pattern to that which emerged from the previous two factors. Those with the highest levels of Irish comprehension were less engaged with the game while those who were in the mid range of the Likert scale tended to be most engaged. These groupings were reflected even in the evaluations of the graphics. Those with the lowest evaluations of the graphics were from the highest language ability group. In contrast, the mid-level pupils, in terms of their understanding of Irish, reported the highest ranking for learning new words/phrases, for the usefulness of the dictionary, and for their overall enjoyment of the game. The ease with which one could cope with the different dialects was also strongly related to one's overall language ability in that those from the highest Irish language ability group reported least difficulty.

Influence of *Attitude Towards Synthetic Voices*

One's attitude towards and acceptance of the TTS synthetic voices in *Digichaint* is strongly related to one's attitude to synthetic voices generally. Those who reported negative attitudes towards synthetic voices generally in the pre-game questionnaire were also negative towards the synthetic voices in *Digichaint,* and vice versa. This is the strongest pattern to emerge from this section of the analysis and one can legitimately claim that one's judgment of *ABAIR* synthetic voices is coloured by one's predisposition towards synthetic speech. It can be said that one's general attitude towards synthetic voices is highly related to one's perception and enjoyment of the *Digichaint* platform and to one's perception of the usefulness of this type of game as an aid to language learning.

5.9.3.3.3 Fáilte go TCD statistical analysis.

This section examines the relationship between data collected from the pre-game questionnaire (background factors) and data from the post-game questionnaire for *Fáilte go TCD*. As with *Digichaint*, the background factors are: gender; school type; frequency of playing computer games; level of understanding of Irish; and general attitude towards synthetic voices.

The post-game items (shown in full in Appendices C3 (in Irish) and E2 (in English)) are summarised as follows:

Table 5.10

Summary of Fáilte go TCD post-game questionnaire items – those referring to TTS aspects are shaded.

Item 1	How would you describe your judgment of the background setting and the graphics in the video?
Item 2	How would you describe your judgment of the body movements of the figures and their alignment to speech?
Item 3	To what extent do the movements of the characters add credibility and clarity to the conversational exchanges?
Item 4	To what extent do you think this type of learning platform (the voices, the graphics and the setting) would help in practising listening comprehension?
Item 5	Would you enjoy using this type of activity to develop your aural Irish skills, should it be available and easily accessible in your school?
Item 6	How motivating do you find this type of activity?
Item 7	Do you think this type of activity would make the learning of Irish more attractive?
Item 8	The overall standard of the Irish used is at about the right level for me.
Item 9	The synthesised voices were sufficiently clear to make the speech **intelligible**.

Item 10	Did you experience particular difficulties with the dialects that are used in the video?
Item 11_1	Please give your opinion on the usefulness of producing graphics with synthesised voices in order to practise aural comprehension.
Item 11_2	Please give your opinion on the **quality** of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here?
Item 11_3	Please give your opinion on the **attractiveness** of the voices.

Table 5.11 below sets out the results of the Kruskal-Wallis analysis for *Fáilte go TCD*. The influence of the five participant background factors are then summarised one by one.

Table 5.11

Results of the Kruskal-Wallis Analysis of the Fáilte go TCD Post-Game Questionnaire Items and Participant Background Factors. Statistically significant differences at $p<0.05$ are denoted in red font with an asterisk. Questionnaire items referring to TTS aspects are shaded.

Item	Gender	School Type	Frequency playing computer games	Level of understanding of Irish	Attitude towards synthetic voices
1	H(1)=0.008, p=0.930	H(2)=5.017, p=0.081	**H(3)=9.928, p=0.019***	**H(4)=18.919, p=0.001***	**H(4)=11.353, p=0.023***
2	H(1)=0.006, p=0.930	H(2)=1.633, p=0.442	H(3)=3.718, p=0.294	H(4)=3.287, p=0.511	H(4)=8.130, p=0.087
3	H(1)=1.357, p=0.244	H(2)=2.327, p=0.312	H(3)=5.852, p=0.119	**H(4)=12.201, p=0.016***	**H(4)=16.570, p=0.002***
4	H(1)=0.218, p=0.641	H(2)=5.181, p=0.075	H(3)=1.460, p=0.692	H(4)=9.478, p=0.050	**H(4)=24.368, p=0.000***

Item	Gender (*continued*)	School Type (*continued*)	Frequency playing computer games (*continued*)	Level of understanding of Irish (*continued*)	Attitude towards synthetic voices (*continued*)
5	H(1)=0.315, p=0. 575	**H(2)=7.138, p=0.028***	H(3)=4.724, p=0.193	H(4)=8.249, p=0.083	**H(4)=23.513, p=0.000***
6	H(1)=0.121, p=0.728	H(2)=5.469, p=0.065	H(3)=3.306, p=0.347	H(4)=3.391, p=0.495	**H(4)=31.689, p=0.000***
7	H(1)=0.820, p=0.365	**H(2)=10.625, p=0.005***	H(3)=5.326, p=0.149	**H(4)=10.958, p=0.027***	**H(4)=29.489, p=0.000***
8	H(1)=0.076, p=0.783	**H(2)=8.059, p=0.018***	H(3)=1.409, p=0.703	**H(4)=13.471, p=0.009***	H(4)=6.831, p=0.145
9	H(1)=2.562, p=0.109	**H(2)=9.589, p=0.008***	H(3)=4.676, p=0.197	**H(4)=17.688, p=0.001***	**H(4)=14.174, p=0.007***
10	**H(1)=4.107 p=0.043***	**H(2)=11.224, p=0.004***	H(3)=0.709, p=0.871	**H(4)=25.683, p=0.000***	**H(4)=11.846, p=0.019***
11_1	H(1)=0.123, p=0.726	H(2)=5.249, p=0.072	**H(3)=12.418, p=0.006***	**H(4)=9.971, p=0.041***	**H(4)=31.088, p=0.000***
11_2	H(1)=0.654, p=0.419	H(2)=3.659, p=0.160	H(3)=7.448, p=0.059	H(4)=5.987, p=0.200	**H(4)=16.508, p=0.002***
11_3	H(1)=3.097, p=0.078	**H(2)=8.762, p=0.013***	H(3)=3.071, p=0.381	**H(4)=11.171, p=0.025***	**H(4)=12.383, p=0.015***

Influence of *Gender*

Gender reached the level of statistical significance for only one item in the post-game questionnaire. This is Item 10 which asks if respondents experienced "difficulty with the dialects" used in the platform. Boys reported significantly less difficulty than girls in dealing with the dialects used ($H(1)=4.107$, $p=0.043$*). The mean ranks for this item were 140.58 for boys and 120.98 for girls.

Influence of *Between-School Differences*

Those from English-medium schools found the *Fáilte go TCD* platform more attractive as a means of engaging with language learning than did those from Gaeltacht schools. Generally those from Gaelscoileanna came in with an intermediate score between the other school types. Gaeltacht school pupils, however, gave a significantly higher rating to the intelligibility of the synthetic voices and to their attractiveness.

Influence of *Frequency of Playing Computer Games*

It may be noted that *Fáilte go TCD*, as presented in this study, does not require manipulation of the figures by the player. It is not surprising, consequently, that the frequency with which respondents play computer games is significantly related to only two of the post-game questionnaire items - namely Items 1 and 11_1.

Frequency of playing computer games was significantly related to the respondents' rating of the graphics in the video ($H(3)=9.928$, $p=0.019$*). Those who never played computer games gave the lowest ranking to the graphics (mean ranks: 'seldom play': 137.31; 'play weekly': 123.33; 'play daily': 115.08; 'never play': 102.92). Those who had some familiarity with playing computer games gave a significantly higher rating to the graphics than those who reported 'never' playing ($p=0.003$*). There is no statistically significant difference between the frequencies with which one played computer games, though the scores show a general tendency for those who play games '*sometimes*', but infrequently,

to give a higher ranking to the graphics. It would appear that those who play computer games very frequently have very high expectations of the standards of the graphics.

The pattern shown in Item 1 repeats itself for Item 11_1. Those who '*never*' played computer games had a significantly lower opinion of the usefulness of such a platform while those who '*seldom*' played gave the idea a significantly higher mean ranking (mean ranks: 'seldom play': 135.12; 'play weekly': 133.58; 'play daily': 117.25; 'never play': 98.76).

Influence of *Level of Understanding of Irish*

Level of understanding of Irish had a significant statistical relationship with eight of the thirteen items on the post-game questionnaire (see Table 5.11). This makes it a very important factor in determining one's overall evaluation of the platform. As with *Digichaint*, respondents are differentiated according to how they placed themselves on a Likert Scale with those with least competence on point 1 and those with highest competence on point 5.

The numbers showing low Irish ability levels (points 1 and 2 on the Likert scale) are very small and it is therefore considered inappropriate to include them in the discussion of results.

One's ability level at Irish has an important influence on one's judgment of the *Fáilte go TCD* platform. Those with intermediate levels of Irish language ability seemed to have a greater attraction to the platform. They had a higher regard for the quality of the graphics and for the alignment of characters' movement and speech. There was a tendency for those with the highest language ability level to be somewhat dismissive of the usefulness of this type of platform. Yet, on the issues of the intelligibility, quality and attractiveness of the synthetic voices it is clear that the higher ability language group gave a much higher ranking to these factors than did those of lower ability. Higher ability groups also

found the comprehension of the different dialects less challenging. Those with lowest overall language ability reported the greatest problems with the intelligibility and attractiveness of the synthetic voices.

Influence of *Attitude towards Synthetic Voices*

Predisposition towards synthetic voices is the background factor which is most strongly related to almost all the post-game questionnaire items. There are just two items where the relationship has not reached statistical significance (Items 2 and 8). The pattern of the relationship remained constant for each of the items with mean rank scores for Group 5 (those who responded that "synthesised voices are *sometimes more suitable than human voices*") higher than those for Group 4 (synthesised voices are '*sometimes suitable*'), which, in turn, were higher than those for Group 3 ('*neutral*'), etc. The precise breakdown of these results is shown in Appendix F2.

A respondent's acceptance of the various aspects of the platform was in direct proportion to his/her predisposition towards synthetic speech generally. Those with a negative predisposition towards the TTS synthetic voices gave a very low ranking to graphics, movement, usefulness of platform and the motivational value of the platform. They also gave a lower ranking to the quality and intelligibility of the synthetic voices as well as expressing difficulties with the dialects used. They had a low ranking for the usefulness of a platform such as this for practising aural comprehension skills and for the quality and attractiveness of the voices. One's rank ordering of all these factors increased according to how one viewed synthetic speech, as reported on the Likert scale in the pre-game questionnaire.

5.9.3.3.4 Taidhgín statistical analysis.

An analysis of the relationship between the background factors elicited from the pre-game questionnaire and items from the post-game questionnaire for *Taidhgín* is given in this section. The pre-game factors, as with *Digichaint* and *Fáilte go TCD*, consist of

gender; school type; frequency of playing computer games; level of understanding of Irish; and attitude towards synthetic voices generally.

The post-game items (which are in Appendices C4 (in Irish) and E3 (in English)) are summarised as follows:

Table 5.12

Summary of Taidhgín post-game questionnaire items - items referring to TTS aspects are shaded.

Item 1	Do you feel the graphic display (the talking monkey) is suitable for this type of game/activity?
Item 2	Do you feel the talking monkey adds to the overall playfulness of the exercise?
Item 3	How would you describe the movements of the talking monkey and their alignment to speech?
Item 4	To what extent do you think this type of learning platform (i.e. the interaction and playfulness) would help in practising conversational Irish?
Item 5	Would you use a virtual conversation partner (like the talking monkey) should it be available and easily accessible in your school?
Item 6	To what extent would you be motivated by this type of activity?
Item 7	Do you think a virtual conversation partner such as Taidhgín would make the learning of Irish more attractive?
Item 8	The overall standard of the Irish used by Taidhgín is at about the right level for me.
Item 9	The synthesised voices were sufficiently clear to make the speech **intelligible**.
Item 10	Did you experience particular difficulties with the dialects that are used in Taidhgín?
Item 11_1	Please give your opinion on the usefulness of the concept of producing a virtual conversation partner who speaks with a synthesised voice in order to practise oral Irish.

Item 11_2	Please give your opinion on the **quality** of the synthesised voices: to what extent do you think the voices are adequate for the type of learning platform presented here?
Item 11_3	Please give your opinion on the **attractiveness** of the voices.

Table 5.13 below sets out the results of the Kruskal-Wallis test with significant statistical differences at $p<0.05$ highlighted and marked with asterisks (*). The discussion that follows focuses on those significant results. See Appendix F3 for detailed results of the Kruskal-Wallis post hoc tests for *Taidhgín*, which are also referred to in the discussion below. See Appendix H for item-by-item discussion of those statistically significant results. The influence of the five participant background factors are summarised one by one below.

Table 5.13

Results of the Kruskal-Wallis Analysis of the Taidhgín Post-Game Questionnaire Items and Participant Background Factors. Statistically significant differences at p<0.05 are denoted in red font with an asterisk. Questionnaire items referring to TTS aspects are shaded.

Item	**Gender**	**School Type**	**Frequency playing computer games**	**Level of understanding of Irish**	**Attitude towards synthetic voices**
1	H(1)=1.424, p=0.233	H(2)=0.070, p=0.966	H(3)=5.478, p=0.140	H(4)=4.178, p=0.382	**H(4)=12.268, p=0.015***
2	H(1)=0.098, p=0.755	H(2)=4.056, p=0.132	H(3)=2.976, p=0.395	H(4)=3.892, p=0.421	H(4)=2.951, p=0.566
3	H(1)=0.208, p=0.648	**H(2)=6.562, p=0.038***	H(3)=0.176, p=0.981	H(4)=1.816, p=0.770	H(4)=8.728, p=0.068

Item	Gender *(continued)*	School Type *(continued)*	Frequency playing computer games *(continued)*	Level of understanding of Irish *(continued)*	Attitude towards synthetic voices *(continued)*
4	**H(1)=5.143, p=0.023***	**H(2)=17.186, p=0.000***	H(3)=2.531, p=0.470	**H(4)=14.067, p=0.007***	H(4)=3.698, p=0.448
5	H(1)=2.351, p=0.125	**H(2)=14.895, p=0.001***	H(3)=6.651, p=0.084	**H(4)=20.580, p=0.000***	H(4)=5.686, p=0.224
6	H(1)=2.460, p=0.117	**H(2)=15.082, p=0.001***	H(3)=2.736, p=0.434	**H(4)=10.411, p=0.034***	**H(4)=10.947, p=0.027***
7	H(1)=0.247, p=0.619	H(2)=4.065, p=0.131	H(3)=2.535, p=0.469	H(4)=5.594, p=0.232	H(4)=5.350, p=0.253
8	H(1)=1.037, p=0.309	H(2)=1.358, p=0.507	H(3)=4.276, p=0.233	H(4)=3.015, p=0.555	**H(4)=11.168, p=0.025***
9	**H(1)=5.135, p=0.023***	H(2)=3.876, p=0.144	**H(3)=10.558, p=0.014***	H(4)=3.925, p=0.416	**H(4)=14.607, p=0.006***
10	**H(1)=7.243, p=0.007***	**H(2)=20.754, p=0.000***	H(3)=5.537, p=0.136	**H(4)=15.908, p=0.003***	**H(4)=15.347, p=0.004***
11_1	H(1)=0.000, p=0.994	H(2)=4.279, p=0.118	H(3)=3.104, p=0.376	H(4)=5.595, p=0.232	H(4)=2.482, p=0.648
11_2	H(1)=1.552, p=0.213	H(2)=1.967, p=0.374	**H(3)=15.218, p=0.002***	H(4)=5.799, p=0.215	H(4)=9.257, p=0.055

Item	Gender *(continued)*	School Type *(continued)*	Frequency playing computer games *(continued)*	Level of understanding of Irish *(continued)*	Attitude towards synthetic voices *(continued)*
11_3	H(1)=0.101, p=0.751	**H(2)=6.408, p=0.041***	**H(3)=8.192, p=0.042***	H(4)=5.799, p=0.206	**H(4)=10.205, p=0.037***

Influence of *Gender*

Gender has a significant relationship with three of the thirteen post-game items. These were Items 4 (H(1)=5.143, p=0.02*), 9 (H(1)=5.135, p=0.02*) and 10 (H(1)=7.243, p=0.01*). Item 4 asked about the "usefulness" of this type of playful interactive platform as an aid to practising conversational Irish. The mean ranking for girls was significantly higher than that for boys, meaning that girls tended to prefer the platform to boys (mean ranks: girls: 120.00; boys: 99.78). Interestingly, boys gave a significantly higher ranking to the clarity and intelligibility of the synthesised voice in Item 9 (mean ranks: boys: 129.24; girls: 108.99). Boys also reported experiencing significantly less difficulty than girls in dealing with the dialects used in *Taidhgín* (mean ranks: boys: 132.50; girls: 107.78).

Influence of *Between-School Differences*

Between-school differences are significant for Items 3, 4, 5, 6, 10 and 11_3 (see Table 5.13) with no significant differences showing for the remaining items. Those from English-medium schools were generally more positive in their attitude to *Taidhgín* as a virtual conversational partner. This probably reflects the lack of opportunities those pupils have for a human conversation partner. This section of the analysis also shows that pupils from Gaeltacht schools had least problems in dealing with dialectal variation which suggests that those variations pose least problems for those with the highest standards of Irish.

Influence of *Frequency of Playing Computer Games*

Frequency of playing computer games had a significant relationship with three post-game items, namely Items 9, 11_2 and 11_3. Those who reported that they 'never' played computer games found the synthetic voices significantly less clear and intelligible than did those who played regularly. This replicates the result got from the corresponding item in *Digichaint* and can most likely be seen as showing that some familiarity with gaming made the experience of dealing with *Taidhgín* more positive which, in turn, led to greater positivity towards the synthetic voices. There is also greater positivity amongst more frequent computer game players towards the "adequacy of the synthesised voice" for a platform such as *Taidhgín* and the highest mean rank scores with regard to the "attractiveness" of the voices also come from the more frequent computer game players.

Influence of *Level of Understanding of Irish*

Level of understanding of Irish had a significant statistical relationship with Items 4, 5, 6 and 10 from the post-game questionnaire. Again, those who rated themselves at the lowest Irish language ability levels (points 1 and 2 on the Likert scale) are being excluded from the present discussion as numbers are very low and it is unsafe to draw any conclusions from the data relating to them.

It would appear that those who have a high language ability and come from a Gaelscoil or from a Gaeltacht school have least interest in engaging with a virtual dialogue partner. They already have human dialogue partners in their immediate environment. Those with an intermediate ability showed significantly more enthusiasm for such a platform, possibly because of lack of opportunity to have human dialogue partners.

Influence of *Attitude towards Synthetic Voices*

One's general attitude towards synthetic voices had an influence on six of the *Taidhgín* post-game items. Significant relationships are noted in Items 1, 6, 8, 9, 10 and 11_3 (see

Table 5.13). In the discussion below, Group 5 is composed of those who responded that synthetic voices are "sometimes more suitable than human voices" in the context of computer games. Group 4 respondents reported that they found synthetic voices "sometimes suitable", Group 3 chose "neutral", Group 2 represents respondents who chose "tolerate synthesised voices but prefer human voices" and Group 1 represents those who responded "I hate synthesised voices".

The results from this item show that those with a positive predisposition towards synthetic voices were also positive about the graphics and found that this type of virtual dialogue partner would be "motivational" for them as a study aid.

Those with a positive predisposition towards synthetic speech were also more likely to have come from English-medium schools. They tended towards reporting that the level of Irish used in the platform was 'suitable' for them (mean ranks: Group 5: 205.50; Group 4: 127.62; Group 2: 116.91; Group 3: 107.86; Group 1: 80.70). This may reflect a lower difficulty level in the language used in *Taidhgín* compared to the other two platforms and a greater familiarity with the limited topics covered.

A positive predisposition towards synthetic voices was also significantly related to judgments on the "clarity and intelligibility of the voice" in *Taidhgín* and those who reported a positive predisposition were significantly less likely to experience "difficulty with the dialects used". It was also associated with a significantly stronger mean ranking on the "attractiveness" of the synthetic voice used in *Taidhgín*.

5.9.3.3.5 Summary of inferential statistical analysis.

There was a remarkably high similarity between the results of the inferential statistical analysis for each of the three platforms.

Gender differences were of very little significance in the opinion ratings of the respondents to any of the platforms.

Gaelscoileanna and Gaeltacht schools formed, for the most part, a coherent group which had a lower opinion of each of the three platforms than that of those from English-medium schools.

A general trend also emerged showing that those who rarely, if ever, played computer games were not very impressed by the platforms. A similar trend applied to those who were very familiar with computer gaming. It is possible that the latter group was comparing the graphics to those of commercially available games. Those with the highest appreciation of the platforms were those who had some general familiarity with gaming and computing.

Those from the highest language ability group gave a lower overall evaluation to the platforms but reported least difficulty in dealing with the different dialects. This strongly suggests that those who saw the platforms as being relevant to themselves in terms of an aid to language acquisition rated them highly.

One of the most significant factors associated with rating the platforms was one's preconceived attitude to synthetic voices generally. Those with negative attitudes tended to give generally lower ratings to most aspects of the platforms.

5.9.3.3.6 A holistic experience? Results of the Spearman's rho analysis.

Spearman's rho is being used to test the relationship between each of the items in post-games questionnaires. It is designed to test, primarily, Hypothesis, that there is a statistically significant relationship between the measured reactions of the respondents to the various non-TTS aspects of the platform and their opinions on the TTS aspects – i.e. the quality and attractiveness of the synthetic voices used in each platform. The initial expectation, based on the literature, is that pupils react to these game-like activities as a holistic experience, and that indeed all aspects may influence one other.

Separate tests were carried out for each of the platforms and details (Spearman's correlation, its significance value, and the sample size that the calculation was based on) are presented in Appendix G.

Discussion on the Spearman's rho Analysis

An inspection of the first table in Appendix G clearly indicates a statistically significant relationship between virtually all the items on the post-game analysis of *Digichaint*. The two items which fall short of statistical significance in their relationship with most other items are Items 9 and 15. In the case of *Fáilte go TCD* all items, with the exception of one, have reached statistical significance either at the $p<0.05$ level or the $p<0.01$ level (see Appendix G2). In the *Taidhgín* platform, there is significant statistical inter-relationship between all items without exception. All correlations are significant at the 0.01 or the 0.05 levels (see Appendix G3). This provides striking support for the initial assumption that the experience of these game-like CALL platforms is holistic.

The H_0 for Hypothesis 5 may be stated as follows: "There is NO statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform."

Post-game questionnaire items relating to intelligibility, quality and attractiveness of the synthetic voices in each of the three platforms have a statistically significant relationship to practically all other post-game questionnaire items. Since practically all of the Spearman's rho correlations have a positive relationship with one other, one is led to reject this H_0 5. This is a highly significant result and offers strong support to one of the core contentions in the present study. This contention is that the acceptability of synthetic speech is context-specific. Morton et al. (2011) found that people accepted synthetic voices when accessing their bank accounts because they felt no human was involved, something that afforded them greater privacy. For information-giving activities a very different finding has been reported, and human voices were found to be preferable (Francis & Nusbaum, 1999, p. 79). It was the initial contention of this study that synthetic speech should be judged in a context-specific way. The present research found that one's level of engagement with the platform as a whole was significantly related to one's judgment of and attitude to the synthetic speech.

The linkage of reactions to the non-TTS and the TTS aspects of these platforms is striking and is underlined by the fact that pupils diverged in terms of their positive or negative attitudes towards the various aspects of the post-game questionnaire, as shown in Appendices E1, E2 and E3. Some regarded graphics and voices as being very attractive while others thought the opposite. This is suggesting that one's attraction to particular language games may be related to one's general attitude towards the use of computers for language learning in general. A lack of appreciation for factors such as graphics, movement and ease of manoeuvring the game will be significantly related to a negative view of synthetic voices. There may be differences between people in their acceptance of virtual environments and, quite possibly, these differences may be associated with factors such as learning styles (Shi, 2011) or empathy towards the use of computers. It is likely that virtual environments may vary in their effectiveness for learners in a way that depends on personality-types.

5.9.4 Implications for hypotheses.

The extent to which the hypotheses stated first in Chapter 3.4 can be supported, is examined in this section. To remind the reader of these hypotheses, Table 5.3 is repeated below. Note that the first hypothesis refers only to *Digichaint*.

Table 5.3

Overview of Hypotheses, as presented first in Chapter 3.4

Hypothesis 1	*Digichaint,* as an interactive game-like platform, will conform to the **criteria** for effective CALL platforms:
1.1	game is task-based with a focus on meaning
1.2	game has an appropriate language difficulty level
1.3	game is easy to manoeuvre
1.4	game is enjoyable
1.5	game has a clear plot
1.6	game has a logical structure
1.7	game is engaging for the learners
1.8	game has graphics that are considered attractive/engaging
1.9	game has legitimate narrative, i.e. credibility for the player/learner
Hypothesis 2	The synthetic speech is readily intelligible for respondents with a high Irish language comprehension ability.
Hypothesis 3	The synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness.
Hypothesis 4	A majority of respondents regard the synthetic voices as being acceptable in the context of the CALL platforms in which they are presented.
Hypothesis 5	There is a statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms (such as plot, ease of manoeuvring characters, legitimacy of narrative and standard of Irish language used) and

	their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform.
Hypothesis 6	There is a statistically significant relationship between respondents' predisposition towards synthetic voices in general and their attitude towards the Irish synthetic voices being tested.

5.9.4.1 Hypothesis 1.

Results here are discussed in terms of the criteria set out first in Table 3.2 (see Chapter 3.3.1.2) and subsequently Table 5.3 of this thesis, which refers to desirable characteristics of good CALL games. Each criterion, in turn, is turn set out in terms of Hypothesis 1 as stated in Chapter 3.4.

Evaluation of these criteria is confined to data pertaining to *Digichaint* since this was the only platform which can be said to have had a task-based problem-solving structure (see Appendix E1 for details of the questionnaire items with responses).

Criterion 1: Games to be task-based with focus on meaning and on use of the target language.

Judgment on Criterion 1 is based on Items 9 and 10 of the post-game questionnaire. The proposition was put to respondents that "I was more focused on the plot of the game than I was on the language structures being used" (Item 9). 54% agreed or strongly agreed that they were mainly focused on the plot while 26% '*disagreed*' or '*strongly disagreed*' with this. 21% remained neutral.

75% '*agreed*' or '*strongly agreed*' that there was a good balance between the playfulness of the game and the opportunities it presented for language learning and only 7% '*disagreed*' or '*strongly disagreed*' with this (Item 10). *Digichaint* is task-based in the sense that it has a plot and an achievable aim. Criterion 1 has been satisfactorily achieved in that learners engaged with the game and agreed that there was a good balance between

plot and language. **Hypothesis 1.1**, that games are to be task-based with a focus on meaning, is supported by these results.

Criterion 2: For this criterion, an evaluation was made of (1) the appropriateness of the language difficulty level and (2) the appropriateness of manoeuvring through the game. A third criterion which could not be evaluated in this study was that the game should afford learners an opportunity to develop metalinguistic awareness.

Item 2 of the questionnaire asked learners to assess the technical ease with which they could manoeuvre through the game. 76% regarded it as being '*easy*' or '*very easy*' to play. Only 5% regarded it as being '*difficult*' or '*very difficult*'. **Hypothesis 1.3**, that games should be easy to manoeuvre, is supported.

In response to Item 3, 66% '*agreed*' or '*strongly agreed*' that they had learned some new words, phrases or grammar points from playing the game.

The difficulty level of the language was regarded as appropriate by 70% of respondents (Item 11). Those who did not agree with this were about equally divided as to whether it was too easy or too difficult. This would suggest that the language difficulty was about appropriate for the general body of 17-year-old learners in Irish second level schools. **Hypothesis 1.2**, that the games have an appropriate language difficulty level, is supported.

Criterion 3: Playful spirit of the game. Enjoyment of the game.

Item 10 showed that 75% thought that there was a good balance in the game between playfulness and opportunities to focus on language structures. **Hypothesis 1.4**, that games are enjoyable, is supported by this result.

Item 17_1 of the questionnaire asked for learners' opinions on the value of the game as a means of practising their language skills. 81% responded positively with only 4% having a negative reaction to it. This suggests that criterion 3 has been very adequately met.

Criterion 4: Clarity of game plot. Helpfulness of hypertext dictionary

Item 4 asked for respondents' opinions on the usefulness of the hypertext dictionary as part of the game. 80% '*agreed*' or '*strongly agreed*' that the hypertext dictionary was a useful addition to the game. Only 6% disagreed with this but it is not possible to state whether this was because they found the dictionary unnecessary or because it was too difficult to manipulate (the language level was well within the competence range of many of them).

77% regarded the game plot as being clear or very clear with only 14% regarding it as being 'unclear' or 'very unclear' (Item 6). **Hypothesis 1.5**, that the game has a clear plot, is supported by this result. 68% regarded the plot as being '*realistic*' or '*reasonably realistic*' (Item 7), with only 12% regarding it as being '*unrealistic*' or '*very unrealistic*'.

Criterion 5: Logical structure of the game. Appropriate clues to successfully manoeuvre through the game.

An insight into the results for this criterion can be gleaned from the responses to Item 5 which sought students' overall opinions of the game as a facility for language learning. It can be assumed that if the game lacked a logical structure or was difficult to manoeuvre, it would not be regarded as pleasant. 69% regarded the game as being '*pleasant*' or '*very pleasant*' while only 8% had the opposite negative view. **Hypothesis 1.6**, which states that the game has a logical structure, is supported here.

Criterion 6: Intensity of Engagement with the Game.

It was not possible to use an objective metric to measure the intensity of engagement with the game. However, Item 8 in the questionnaire asked respondents to react to the proposition that they remained concentrated while they played the game. 66% '*agreed*' or '*strongly agreed*' with this while 17% '*disagreed*' or '*strongly disagreed*'. Personal observation of the learners playing the game would confirm this high degree of concentration for its duration. They completed the games within a short time of one other, which would suggest that they did not spend much time daydreaming! A record of each user's moves and choices in the game was made but it was considered that a detailed analysis of this would not be particularly useful for the present purposes. The level of engagement observed was high when one considers that the game took more than 25 minutes to complete. **Hypothesis 1.7**, that the game is engaging for learners, is supported by this result.

Criterion 7: Appropriateness of visual and aural material in the game

The aural aspect or synthetic voice will be dealt with separately under Hypothesis 2 below. The visual appropriateness of the game is gleaned from responses to Item 1 which asked for respondents' opinions on the quality of the graphics. 60% regarded the graphics as being '*good*' or '*very good*' while only 14% regarded them as being 'weak' or 'very weak'. 26% classified them as being 'reasonable'. **Hypothesis 1.8**, which states that the graphics in the game are considered attractive/engaging, is supported by this result.

Criterion 8: Legitimacy of game narrative

The legitimacy of the game narrative in this case can be regarded as a combination of the legitimacy of the plot (Item 7) and the appropriateness of the language level (Item 11). In both cases there was a strong positive reaction towards the game. **Hypothesis 1.9**, that the game has a legitimate narrative, is supported by these results.

Criterion 9: Clarity of Screen Layout

Reaction to this criterion is gleaned from responses to questionnaire items on the layout of the graphics and the ease of manoeuvring through the game (Items 1 and 2). There was a strong positive reaction to both of these items.

Summary

An analysis of the data obtained from the pupils shows that each subsection of **Hypothesis 1**, (which states that "*Digichaint,* as an interactive game-like platform, will conform to the criteria for effective CALL platforms", as set out in Chapter 3.3.1.2) is supported.

5.9.4.2 Hypothesis 2.

H_0 2: The synthesised speech is NOT readily intelligible for respondents with a high Irish language comprehension ability.

H2 has already got support from teacher data, as shown in Chapter 4.9. It is further supported by data from the post-game descriptive results (see Section 5.9.3.2) for all three platforms, in that a majority rated the synthetic speech as being 'intelligible'.

Further analysis was carried out through the inferential statistics into the proposition that one's level of Irish is related to one's view on the intelligibility of the synthetic speech. It was postulated that School-Type (ST) and self-assessed Level of Understanding of Irish (LUI) would act as a proxy for pupils' language comprehension ability level and it was suggested that those from Gaeltacht schools, or Gaelscoileanna would be more likely to agree that the synthetic speech is readily intelligible. It was speculated that the same would apply to those who gave themselves the highest rating on the Likert scale for their Irish Language comprehension ability (LUI).

Digichaint

Results from the Kruskal-Wallis Test show that there was no significant difference (H(2)=1.621, p=0.445) between school types in their view that "synthesised voices were clear enough to make the speech intelligible" (Item 12, see Appendix E1). Likewise, there were no between-school differences in the pupils' responses to Item 17_2 which asked for pupils' opinions on the quality and adequacy of the synthetic speech (H(2)=3.551, p=0.169). Items 15 and 16 refer to the atmosphere which the synthetic voices created for computer games and to the suitability of the computer-generated voices for *Digichaint*. No between-school differences were found.

A between-school difference was found in the case of Item 14, which asked pupils to react to the statement that they "found it no more difficult to understand the computer-generated voice than they would if natural voices were used" (H(2)=6.739, p=0.034*). Unexpectedly, it was shown that it was those from English-medium schools who were most strongly in agreement with this statement. At first, this may seem surprising but it may point to a lack of experience of language use by those in the English-medium schools, whereas those from Gaelscoileanna and Gaeltacht schools have a greater experience of dealing with Irish speakers either in school or in their communities, and would be more cognisant of the difference between natural voices and synthetic voices.

Overall, there was little or no between-school differences in pupils' reaction to the synthetic voices. It may be that most pupils had a high enough standard of Irish to make the speech readily intelligible to them.

Summary: School-Type

ST x Item 12: H(2)=1.621, p=0.445. Not significant ($p < 0.05$).

ST x Item 14: H(2)=6.739, p=0.034*. Significant ($p < 0.05$).

ST x Item 15: H(2)=0.939, p=0.625. Not significant ($p < 0.05$).

ST x Item 16: H(2)=5.001, p=0.082. Not significant ($p < 0.05$).

ST x Item 17_2: H(2)=3.551, p=0.169. Not significant ($p < 0.05$).

Similar tests were carried out on the same items (Items 12, 14, 15, 16, 17_2) and again, no differences were found between those who gave themselves a relatively low ranking in their language comprehension ability (LUI) and those who gave themselves a high ranking. It must be remembered, however, that very few gave themselves a ranking of 1 or 2 on the Likert scale for their level of understanding of Irish (only 5 pupils out of 250) and these were excluded from the analysis because of extremely small numbers.

Summary: Level of Understanding of Irish

LUI x Item 12: $H(4)=1.064$, $p=0.900$. Not significant ($p < 0.05$).

LUI x Item 14: $H(4)=3.103$, $p=0.541$. Not significant ($p < 0.05$).

LUI x Item 15: $H(4)=6.166$, $p=0.187$. Not significant ($p < 0.05$).

LUI x Item 16: $H(4)=6.237$, $p=0.182$. Not significant ($p < 0.05$).

LUI x Item 17_2: $H(4)=9.165$, $p=0.057$. Not significant ($p < 0.05$).

Fáilte go TCD

Items 9 and 11_2 ('intelligibility' and 'quality' of synthetic voices) of the *Fáilte go TCD* post-game questionnaire were correlated with School Type (ST) and Level of Understanding of Irish (LUI). ST and LUI are taken as a proxy for the language comprehension ability of pupils. It was anticipated that those from Gaeltacht schools and Gaelscoileanna, as well as those who gave themselves a high ranking on the comprehension ability scale, would have a significantly higher mean rank score for Items 9 and 11_2 than would those from English-medium schools.

Results of the Kruskal-Wallis test show that for Item 9 those from Gaeltacht schools were more likely to find that "the voices were sufficiently clear to make the speech intelligible". However, there were no significant differences between school types in their reactions to Item 11_2 which stated that "the quality of the synthesised voices was adequate for this type of language learning platform".

ST x Item 9: H(2)=9.589, p=0.008 *. Significant ($p < 0.05$).
ST x Item 11_2: H(2)=3.659, p=0.160. Not significant ($p < 0.05$).

Exactly the same pattern holds for LUI where those with the highest level of Irish were more likely to find the speech intelligible (Item 9) but there were no significant differences between groups in relation to their reactions to the quality and adequacy of the synthesised voices for this type of learning platform (Item 11_2).

LUI x Item 9: H(4) = 17.688, p=0.001 *. Significant ($p < 0.05$).
LUI x Item 11_2: H(4)=5.987, p=0.200. Not significant ($p < 0.05$).

Taidhgín

As with the two previous platforms, School Type (ST) and Level of Understanding of Irish (LUI) were correlated with Items 9 ("the synthesised voices were sufficiently clear to make the speech intelligible") and 11_2 ("quality of voices / adequacy for this type of learning platform").

ST x Item 9: H(2)=3.876, p=0.144. Not significant ($p < 0.05$)
ST x Item 11_2: H(2)=1.967, p=0.374. Not significant ($p < 0.05$).

This shows that there were no significant differences between school types in their judgment of the intelligibility of the synthetic voices.

LUI x Item 9: H(4)=3.925, p = 0.416. Not significant ($p < 0.05$).
LUI x Item 11_2: H(4)=0.215, p = 0.215. Not significant ($p < 0.05$).

Similarly, this shows that there were no differences between respondents at different levels of understanding of Irish and their judgment on the intelligibility of the synthetic voices.

Summary

Hypothesis 2 is supported by reference to the descriptive statistics where a majority of respondents (both teachers and pupils) rated the synthetic voices as being 'sufficiently clear to make the speech intelligible'.

On further analysis of the data, it emerged that there were few significant differences between school types (ST) or self-rated level of understanding of Irish (LUI) and questionnaire items relating to the intelligibility, quality and attractiveness of the synthetic voices. It may be that those with a high Irish language ability are more familiar with Irish spoken with natural voices and consequently have higher expectations of the speech output. On the other hand, learners, or those with a lower standard of comprehension of the synthetic speech used in the games may be more tolerant of the imperfections of the voices. This is in line with the Kang et al. (2008) findings that learners were more tolerant of synthetic speech than were native speakers.

5.9.4.3 Hypothesis 3.

Hypothesis 3, that the synthetic voices are regarded by a majority of respondents as having a high quality and attractiveness, is partially supported, in that a majority found the synthetic speech to be intelligible and to be of high quality. This suggests that the synthetic speech from *ABAIR* is at a point in its development where it can make a positive contribution to the development of CALL platforms for the teaching/learning of Irish. It can be noted, however, that the majority of respondents did not regard the synthetic voices as being "attractive", suggesting that they would prefer, if possible, to have a human voice. These results are in line with almost all previous research on the relative 'attractiveness' of synthetic speech (Andraszewicz, Yamagishi, & King, 2011; Francis & Nusbaum, 1999).

5.9.4.4 Hypothesis 4.

Hypothesis 4 states that "a majority of respondents regard the synthetic voices as being acceptable" in the context of the CALL platforms in which they are presented. In relation to pupil data, Item 16 ("I found the computer-generated voices suitable for this computer game") of the post-game questionnaire for *Digichaint* (see Appendix E1) refers directly to this hypothesis. 57% of the pupils 'agreed' or 'strongly agreed' with the statement and only 15% 'disagreed' or 'strongly disagreed'. One can also refer here to Item 17_2 which asks pupils to give their opinion on the quality of the voices. 62% saw the voices as having 'adequate' or 'totally adequate' quality while only 21% had the opposite opinion.

For Item 11_2 in the *Fáilte go TCD* post-game questionnaire just over 50% found that the quality of the voices was '*adequate*' or '*totally adequate*' for the platform (Item 11.2, Appendix E2). The figure for the corresponding item for the *Taidhgín* platform was 73% (Item 11_2, Appendix E3).

When the data from both the teachers' (see Chapter 4.9) and pupils' responses are combined one can claim quite a strong level of support for **Hypothesis 4**.

5.9.4.5 Hypothesis 5.

The H_0 for Hypothesis 5 may be stated as follows: "There is NO statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms and their opinions on the intelligibility, quality and attractiveness of the synthetic voices used in each platform."

This H_0 was tested by the use of Spearman's Rank Correlation Coefficient, which tested for significant correlations between each of the post-game questionnaire items. Full results of these tests are given in Appendix G. A discussion on the analysis is given in Section 5.9.3.3.6. It is noted that there is a highly significant relationship between all post-game questionnaire items. This gives strong support to the proposition that one experiences a CALL platform as a holistic experience and that one's opinion on any

particular aspect of the platform is influenced by one's opinion of all other aspects of the same platform. Of particular significance here is the proposition that one's judgment of the synthetic speech is influenced by the context in which it arises and, therefore, the use of synthetic speech in CALL platforms is conditional on the overall quality of those platforms.

5.9.4.6 Hypothesis 6.

H_0 6: *There is NO significant relationship between respondents' predisposition towards synthesised voices in general and their attitude towards the Irish synthesised voices being tested.*

This H_0 is tested in the case of each of the three platforms by reference to the correlations between Attitude towards Synthetic Voice (ASV) and Items 12, 14, 15, 16, 17_2 and 17_3 for *Digichaint*; Items 9, 11_2 and 11_3 for *Fáilte go TCD* and *Taidhgín*, respectively.

Digichaint

Table 5.9 shows a significant relationship between predisposition towards synthetic voices and Item 12 ("intelligibility of synthesised voices"); Item 15 (the "atmosphere" synthesised voices give to games); and Item 17_2 (the "quality" and "adequacy" of the synthesised voices for games such as *Digichaint*). The null hypothesis can be rejected in the case of each of these items and a significant relationship between predisposition towards synthetic voices and the above factors is established. This shows that, at the present stage in the development of synthetic voices, most people have experienced computer-generated voices in some context and this experience colours their attitude to each new synthetic voice they hear.

Summary: Attitude towards Synthetic Voice

ASV x Item 12: H(4)=12.952, p=0.012*. Reject H_0 ($p < 0.05$).

ASV x Item 15: H(4)=11.785, p=0.019*. Reject H_0 ($p < 0.05$).

ASV x Item 17_2: H(4)=10.693, p=0.030*. Reject H_0 ($p < 0.05$).

The relationships between the remaining three items fell short of significance. These items were Item 14 (synthesised voices as easy to understand as natural voices); Item 16 (suitability of synthesised voices for this game); and Item 17_3 (attractiveness of the voices).

ASV x Item 14: H(4)=7.342, p=0.119. Accept H_0 ($p < 0.05$).

ASV x Item 16: H(4)=8.504, p=0.075. Accept H_0 ($p < 0.05$).

ASV x Item 17_3: H(4)=2.455, p=0.653. Accept H_0 ($p < 0.05$).

A general acceptance of synthetic voices influenced respondents' rating of *ABAIR* synthetic voices in terms of their 'intelligibility' and 'adequacy'. This, however, did not extend to influencing respondents' opinion on 'naturalness' or 'attractiveness'.

Fáilte go TCD

The H_0 is tested by reference to Items 9, 11_2 and 11_3 on the post-game questionnaire. Those with a positive predisposition towards synthetic voices in general had a significantly higher opinion of the quality and intelligibility of the synthetic voices (Item 9). Respondents with a positive predisposition towards the synthetic voices were also significantly more likely to give a higher rating to the quality and adequacy of the synthetic voices for *Fáilte go TCD* (Item 11_2). The same pattern held for respondents' opinions on the attractiveness of the synthetic voices (Item 11_3).

ASV x Item 9: H(4) = 14.174, p=0.007 *. Reject H_0 ($p < 0.05$).
ASV x Item 11_2: H(4) = 16.508, p=0.002 *. Reject H_0 ($p < 0.05$).
ASV x Item 11_3: H(4) = 12.383, p=0.015 *. Reject H_0 ($p < 0.05$).

This is the clearest example of where one's prior attitude towards synthetic voices has a strong influence on one's opinion of the synthetic voices being evaluated.

Taidhgín

The H_0 is tested by reference to post-game questionnaire Items 9 ("The synthesised voices were sufficiently clear to make the speech intelligible"), 11_2 ("quality / adequacy of synthesised voices") and 11_3 ("attractiveness of the voices").

The H_0 is rejected in the case of Item 9 and Item 11_3. The correlation falls just short of statistical significance, however, in the case of Item 11_2.

ASV x Item 9: H(4)=14.607, p = 0.006*. Reject H_0 ($p < 0.05$).
ASV x Item 11_2: H(4)=9.257, p = 0.055. Accept H_0 ($p < 0.05$).
ASV x Item 11_3: H(4)=10.205, p = 0.037*. Reject H_0 ($p < 0.05$).

These analyses show that in the case of the *Taidhgín* platform, one's predisposition towards synthetic voices had a significant bearing on one's judgment of the synthetic voice under consideration.

Summary

Overall, based on data from across the three platforms, there is strong support for **Hypothesis 6** (and the rejection of H_0) that respondents' predisposition towards synthetic speech influences their judgment of the present *ABAIR* synthetic voices.

5.9.5 Results: summary and discussion.

This section is divided into two parts – Section 5.9.5.1, which gives a summary and discussion of the descriptive results and Section 5.9.5.2, which gives a summary and discussion of the inferential statistical results.

5.9.5.1 Descriptive results: summary and discussion.

This section has given a breakdown of the results of pupils' responses to questionnaire items relating to each of the three platforms. These results are first given in terms of basic descriptive statistics in Section 5.9.3.

A major aim of this analysis was to examine the intelligibility, quality and attractiveness of the synthetic voices produced by *ABAIR* in the context of each of the three newly developed CALL platforms. This section aimed to describe the acceptability, or otherwise, of each of the platforms in terms of their design, graphics, plot and the level of engagement they stimulated amongst the respondents under 'post-game descriptive results (non-TTS and TTS aspects)', Section 5.9.3.2. A separate, though inter-related analysis was made of the pupils' responses to the synthetic voices under 'TTS descriptive results for the three platforms' presented in Section 5.9.3.2.4.

The data for the analysis were acquired by means of a 5-point Likert scale and overall one can say that each aspect of the platforms enquired into had a relatively high acceptance rate. If one accepts points 4 or 5 on the 5-point Likert scale as representing a strong positivity, then there was an over 70% positive reaction to practically all of the features of the platform which were being examined.

The variations in the responses across the platforms were relatively small, i.e. the graphics, the plot, the level of engagement, etc. got a relatively similar response rating from the pupils. Overall, one may claim that the criteria set out in Chapter 3.3.1.2 for the development of suitable game-like CALL platforms has been satisfied and that the platforms as developed have been judged by pupils to be 'fit-for-purpose'.

Analysis was made of pupils' reactions to the intelligibility, quality and attractiveness of the synthetic voices and the results are presented in Table 5.7 above in terms of the percentage who selected points 4 or 5 on the Likert scale. One can say that there was an overall positive reaction to the voices, though they received a lower rating than did the reactions to the other elements of the platforms. One may note also that there were considerable variations in the responses to the questionnaires. This was considered encouraging as it showed an honesty amongst pupils in their responses, thereby giving greater reliability to the tables. The pupils did not take the safe option of going for point 3 (or neutral position) nor could they be said to have given what they may have perceived as the 'polite' or 'expected' response.

Support for Hypotheses 1, 3 and 4 is provided by this descriptive aspect of the analysis. In the following section, (Section 5.9.3.3), the analysis of the results is taken to a more inferential level in order to account for the variation in the post-game responses given by pupils.

5.9.5.2 Inferential results: summary and discussion.

Section 5.9.3.3 of this chapter examined pupils' responses to the *ABAIR* synthetic speech in the context of the three platforms in which they were presented. Since there was a broad spectrum of opinion on most of the questions being explored, an analysis was carried out on the relationship between key background characteristics pertaining to the respondents and their attitude to various features of the platform. The key issue was the pupils' reaction to the quality of the synthetic speech in the context of each of the three platforms.

As was the case with teacher responses discussed in Chapter 4, the key finding was that the synthetic speech, though not perfect, is fit-for-purpose for game-like CALL platforms. Teachers, who can be thought of as having a high comprehension ability level in Irish, gave a relatively high rating to the quality and intelligibility level of the speech. They found the speech highly intelligible as is evidenced by their ability to transcribe it. One must keep in mind here that the speech samples which they were asked to transcribe

were those which were anticipated to have particular difficulties and were not chosen as being representative of the overall corpus of the *ABAIR* output. Similarly, those from Gaeltacht schools gave a significantly higher rating to the intelligibility of the speech than did those from English-medium schools.

Using learners to assess the quality of synthetic speech poses certain problems in that a low rating could indicate a lack of intelligibility in the speech, but it could also be indicative of a poorer comprehension level in the learner. A low level of comprehension could be related to the learner rather than to the speech quality. This leaves a complex relationship between pupils' backgrounds and their attitudes towards the platforms in general and towards their opinions of the synthetic speech in particular. In the context of *Digichaint*, respondents from English-medium schools, whose comprehension ability may be seen as lower than that of pupils from either Gaelscoileanna or Gaeltacht schools, gave a higher rating to the platform and to the quality of the synthetic speech than did their Gaelscoil/Gaeltacht school counterparts. This is being interpreted as meaning that the English-medium school pupils, having less experience of natural voices in Irish, had fewer natural voices to compare with and were more tolerant towards the imperfections of the synthetic speech. This agrees with the Kang et al. (2008) contention that learners are less sensitive to differences between synthetic speech and natural speech.

'Clarity' and 'intelligibility', though clearly related, may be distinguished from each other. Those from Gaeltacht schools were more in agreement with the proposition that the voices were sufficiently clear to make the speech 'intelligible' in the context of all three platforms. They also had less difficulty in dealing with dialectal variation in the voices. While they regarded the speech as being 'intelligible', they gave a lower rating to its quality and attractiveness. This finding for pupils mirrors to some extent the teachers' reactions to the synthetic speech, where their transcription performance, 'intelligibility' scores exceeded their scores on the 'ease' of transcribing the sentences. Overall, one is tempted to say that both groups find the speech 'intelligible' but less impressive in terms of 'clarity'.

One may conclude that those from English-medium schools were prepared to give the synthetic speech the benefit of the doubt and express a positive view towards it. Those who have greater exposure to natural voices speaking Irish had a greater intolerance for the imperfections of the synthetic speech, even though the speech was readily intelligible to them.

It is interesting to note that English-medium schools gave a higher rating to many aspects of the platforms than did Gaeltacht schools or Gaelscoileanna. The likely interpretation of this finding is that, as their Irish comprehension abilities are ower, they saw the platforms as having greater relevance to themselves. Those who felt that the platforms afforded them little by way of opportunity to learn were less likely to engage. Those in English-medium schools found the language used to be equal to or higher than their existing comprehension ability levels. This put them in the position where there was a greater possibility that they were going to acquire more language by their participation in the activities pertaining to the platforms. Those from Gaeltacht schools and Gaelscoileanna found the language level below that of their own comprehension level and were therefore less likely to benefit from the platforms. These findings support the proposition that CALL materials should be at a level which makes them clearly relevant to the learners and should be at an appropriate level of challenge (Mitchell & Savill-Smith, 2004).

Relatively little academic research has been carried out on the regularity with which teenagers play computer games. Lenhart, Kahne, Middaugh, Rankin Macgill, Evans and Vitak (2008) claim that 97% of American teenagers play computer games and that 50% of those questioned played games "yesterday". This research has not been replicated in Ireland but the present sample showed relatively few playing games regularly. This is more in line with the results of the Kennedy et al. (2007) Australian study who found that less than 15% of first year university students played computer games on a daily basis. The majority played once a month or even more seldom.

It was found that some familiarity with computer game playing was a significant contributor to one's rating of the present platforms. Those who never played computer

games placed a low rating on the value of the platforms and of the synthetic voices. It would appear that the place of computers in today's society may have a significant influence on the value which will be placed on CALL games in the future. This, of course, is a rapidly changing scene.

The predispositions of respondents to synthetic speech in general had a relatively strong impact on their views on the present synthetic speech. Those with negative views on synthetic speech gave a low rating to the present speech and vice versa. This is in line with the Cryer and Home (2009) findings where they state "Talking Book users' acceptance of synthetic speech is likely to depend on previous experience and personal preference". Predisposition towards synthetic speech is clearly based on one's past experience of it. As the quality of synthetic speech improves one may expect that people will have a more positive view of it and become more accepting of it in new situations such as CALL games.

It was postulated that each of the three platforms would be judged as a unit in itself and that the rating of each of its parts would be influenced by the respondents' overall impression of the platform. The rating given to the synthetic speech would be related to most other features of the platforms. In Section 5.9.3.3, correlation tables are given for the relationship between different aspects of each of the three platforms. There was a significant statistical correlation between almost all of the features of the platforms. This indicates that learners' views on most aspects of the CALL platform are dependent on their overall engagement with the activity. This adds support to the claim that evaluation of synthetic speech must be done in the context where the speech is intended to be used (Chapelle, 2001; Cryer & Home, 2009; Handley, 2005; Huang, Acero, & Hon, 2001). It further suggests that one's opinion of the synthetic speech may not be consistent. One may give it a positive rating if one is experiencing a high level of engagement with the context in which the synthetic speech is being used.

The findings on the attitudes of pupils to the three platforms developed for the present project are encouraging. Each of the platforms was most attractive to those who had

intermediate levels of Irish and who could see clear benefits in using this type of digital media to augment their learning. Platforms need to have a clear relevance for their users. Further research is needed before one can speculate as to the relative importance of features such as plot, flow, and narrative when compared to learners' interpretation of the usefulness of the platform as a tool for their own learning. Those with a high level of Irish were less likely to be enthusiastic about the future development of these platforms because they were somewhat irrelevant to them as learning supports. The platforms in the present project are intended for learners who have relatively little comprehensible input (Long, 1985, 1996) by way of native speakers of Irish. There will be further discussion on this issue and on the future directions of CALL in Chapter 6.

Chapter 6: Conclusions and Future Outlook

This thesis set out to develop CALL platforms for Irish which exploited the recently developed Irish synthetic voices of the *ABAIR* project. Along with the platform development, it set out to evaluate the platforms both in a general way and with specific attention to the suitability of the synthetic voices for such educational applications. This entailed eliciting judgments from teachers and school pupils on a wide range of platform characteristics as well as on the intelligibility, quality and attractiveness of the *ABAIR* voices.

The use of TTS synthesis is a relatively new development in the broader field of CALL and one which has not, to date, gained widespread acceptance (Gupta & Schulze, 2012). The question posed for the present study is similar to that posed by Kılıçkaya (2006), amongst others. Is the speech authentic enough, of a sufficiently high quality, or 'fit-for-purpose', for use in specific game-like language learning platforms? It is important to note that this question is being posed in the context of an endangered minority language which nonetheless has large numbers of school learners, most of whom do not have access to native speaker models of the language. It is an initial contention that the acceptability of the platforms and the possible role of such interactive games in language teaching/learning are likely to be quite different from what pertains to widely spoken world languages.

A review of literature on the topic shows that, to date, synthetic speech has been used mainly to help learners with individual language skills such as aids to pronunciation, reading assistants, facilitators of dictation exercises, tests of listening comprehension or talking dictionaries. If, however, TTS synthesis is found 'fit-for-purpose' to interact with the learner in a non-linear fashion we then get near to personalised CALL platforms where the learning can be supported by a rich variety of text inputs (Conlan, O'Keeffe, Brady, & Wade, 2007). There can be a far greater degree of interactivity between learner and machine where there is at least a semblance of the negotiation of meaning between

an avatar of some type and a learner. This concept of negotiation of meaning is a central theoretical underpinning for communicative approaches to language teaching/learning (Foster & Snyder Ohta, 2005; Little, 1994).

In terms of CALL, the negotiation of meaning would involve an avatar being capable of reacting to a learner's input in a meaningful way so that the learner can make the appropriate adjustments to his/her input in order to make grammatically and contextually appropriate inputs. TTS synthesis would be central to the development of such learning/teaching platforms and this would open up an entirely new vista in the direction of CALL development. Speech recognition facilities would have an equally important contribution to make to the development of these platforms but since speech recognition technology can still be problematic even for native speakers, and are non-existent for Irish, it is difficult to envisage it having anything more than a very restricted role with the imperfect pronunciations of a learner. Ehsani and Knodt (1998) referred to the difficulties attached to speech recognition systems for CALL and this explains why they have not made any major impact in the field of CALL in the intervening years. The complex cognitive processes associated with human speech recognition cannot be easily reproduced digitally. Bergström (2007), for example, reports using speech recognition technology for improving pronunciation for learners of Swedish, but there has been no realistic attempt at using the technology in any kind of communicative format. For now, at least, speech synthesis offers greater possibilities for the development of CALL than does speech recognition technology. Evidence of adequacy of a TTS synthesis system for use in interactive game-type language learning platforms would add considerably to extending the possibilities for the direction of CALL development. While the research was conducted in the context of Irish pupils in the 16 – 17 year old age group who are learning the Irish language as a compulsory subject in schools in the Republic of Ireland, the results of the study may be generalised to many other language learning situations. The present chapter summarises the fundamental language acquisition theory on which the research is based, the social context in which the research was conducted, the development of the research materials, the testing and analysis process. Finally, proposals

will be made on how the future development of CALL will need to incorporate this multiplicity of factors.

6.1 Background Assumptions of the Research

The communicative approach to the language learning process is the prevailing theory of the present time and has in fact been the predominant theory since the 1970s. Task-based language learning (TBLL) is the most specific branch of communicative approaches and is the underpinning theoretical background to the present work. Ellis (2005) saw task-based approaches as providing "opportunities for students to attend to and perform pragmatic meaning" (p. 3). There is a further assumption that learners acquire language in relatively predictable stages provided they have exposure to suitable "comprehensible input" (Krashen, 1987). Skehan (1998) went as far as to claim that instruction had no effect on language learning. What is important are interactional adjustments such as those involved in the negotiation of meaning. These cause modifications in both the linguistic input given to the learner and in turn, the learner's linguistic output is also modified. This negotiation and adjustment are central to second language acquisition (Foster & Snyder Ohta, 2005).

In the context of TBLL theory, it was considered necessary to construct a platform, or platforms, where learners would have the opportunity to interact with the computer in a meaningful way in order to achieve some given aim. One needed to aim to achieve an active learning scenario where the learner has "control of the context, pace and scope of their learning experience" (Dagger, Wade, & Conlan, 2005, p. 11).

Gaming, as an approach to second language acquisition, is beginning to become noticed in CALL literature. Most of the games reported involved the collaboration of a native speaker and a learner in some kind of cooperative enterprise in order to achieve the aims of the game. Avatars are frequently used as an intermediary between the learner and

native speaker with communication frequently being written rather than spoken. A comprehensive review of these studies is available in Peterson (2010). Research on the effectiveness of these projects tends to be of a subjective, if not to say anecdotal, nature in that it reports on relatively small groups of university students over a short period of time. Hubbard (2005) reported that only 30% of studies had 31 or more subjects.

In the context of Irish language learners in schools, it is not realistic to try to promote some kind of digitally mediated collaborative venture between native and non-native speakers. In Chapter 2.2 of this thesis the social context of the Irish language is explored. The language is in a somewhat precarious position, where there are no longer any native monolingual speakers. All teenagers in Gaeltacht areas have a command of the English language which is on a par with teenagers from any other geographical area of Ireland and they would therefore derive no linguistic benefit from digitally mediated collaborating exercises. Stand-alone platforms offer the only realistic possibility of digitally-mediated personalised language learning opportunities for learners of Irish. There is a real paucity of such platforms, purpose-built from the ground up, for any language. For the purposes of the present project it was necessary to design platforms from the ground up so that they would be suitable environments in which to test the TTS synthetic speech output of *ABAIR*.

6.2 The Development of the Research Platforms

An essential challenge of the present project was to set up a two way learner-avatar communication system where the synthetic speech formed a core part of this communication. As there were no previous CALL platforms available for Irish that made use of synthetic voices, a major part of this study involved the development of three individual platforms, with integrated synthetic voices, each of which had its own unique characteristics. The aim ultimately will be to create a platform that would bring together the most successful features in each of the three platforms, combining them in a single

task-based scenario, which would engage the learner in a collaborative interaction with an avatar/agent. In the present three platforms, while the focus remained on the task, due consideration was given to the meta-linguistic support required by the learner.

An ideal platform needs to resemble the collaborative human-agent communication model proposed by Hanna and Richards (2013) but instead of multiple individuals cooperating in a virtual environment, as is proposed in their research, the negotiation can be left to one single individual learner who can be directed and encouraged by a virtual agent. This is a type of Intelligent Tutoring System (Corbett, Koedinger, & Anderson, 1997) where the avatar can both encourage the learner to negotiate particular pathways through the system and give corrective feedback resembling that which would be given by a human tutor. This type of platform is referred to here as a 'game', even if it does not strictly conform to all the criteria normally associated with computer games. A general review of computer games and of the usefulness of serious games as part of a CALL system serves to underline the scarcity of such platforms and further reveals that few have been tested by more than small numbers of evaluators judging prototypes.

In order to emulate the real-life situation in which the target language is likely to be used, a CALL platform should ideally have tasks which have a real-world significance, have a focus on meaning, and be able to engage the learners' interests (Dickey, 2011). The cognitive load, or the amount of cognitive effort it takes from the learner, should be at an appropriate level. Screen layout design should be such that it does not add to cognitive load. Its format should be as familiar as possible, the screen should avoid having too much material, and extraneous material should be excluded. The language level should be challenging but not beyond the scope of the learner. Lim and Chai (2008) drew attention to the fact that game playing is a holistic experience and that one's judgment of any one aspect of the game is influenced by one's impression of the other aspects of the game. It seems likely then that one's view of the design features of the platform in general would colour one's view of the synthetic speech. This amounts to a claim that judgments made of synthetic speech may not be consistent but are likely to vary with context. It was thus

necessary in this study to pay a lot of attention to detail when considering the design of the platform on which the TTS synthetic speech could be tested.

As pointed out by Ward (2001), the development of the 'ideal' platform will necessitate a team of people including language instructors, graphic designers and computer programmers. The construction of a single 'ideal' platform, which would conform to all the criteria for a game-like CALL application, would be extremely complex and beyond the resources of the present project. Fortunately, two existing programs became available to me from within Trinity College, Dublin, which were developed for purposes other than CALL platforms for the Irish language. Each of these encapsulated features which one would ideally wish to have in a CALL platform. It became possible to adapt and to further develop each of them, as separate platforms, to suit the purposes of the present project.

The first of the platforms developed, ***"Digichaint"***, was based on the Language Trap, developed by Dr. Neil Peirce and Prof. Vincent Wade, of the Knowledge and Data Engineering Group, within the School of Computer Science and Statistics (Peirce & Wade, 2010). This, in turn, was part of the ELEKTRA project, a European interdisciplinary research project supported by the priority "Information Society Technologies" (IST) within the sixth framework programme, which aims to promote technology enhanced learning in a variety of disciplines. The architecture of the Language Trap was such that it offered personalised learning in four different forms:

(a) adaptive dialogue difficulty
(b) performance feedback
(c) motivational support
(d) metacognitive hints

The learners could take the persona of different characters from within the environment and interact with the other characters from the perspective of their chosen character.

The second platform, titled ***"Fáilte go TCD"***, was derived from the Metropolis Project (funded by Science Foundation Ireland) as developed by a team including Dr. Cathy Ennis and led by Prof. Carol O'Sullivan of the Graphics, Vision and Visualisation Group in the School of Computer Science and Statistics, TCD. The Metropolis Project aims to create a simulated realistic populated virtual environment where the user can move around and experience total immersion in a computer-generated virtual Dublin (O'Sullivan & Ennis, 2011). The particular segment chosen for *Fáilte go TCD* was located in the Front Square of Trinity College and involved graphically realistic-looking people conversing. It did not originally have an interactive task-based game-like plot or storyline or any voice accompaniment as such. Thus, I was free to construct a localised dialogue which would fit with the realistic setting and body movements of the characters.

The *Digichaint* platform offered many of the features which were adduced as being important for an interactive CALL platform as set out in Chapter 2.5.7:

(a) *Digichaint* was task-based with focus on meaning
(b) It promoted metalinguistic awareness and had a playful spirit and clear plot
(c) Additional metalinguistic information was provided by way of hypertext
(d) It was multi-modal in the sense that it had visual, auditory and manipulative dimensions

The *Fáilte go TCD* platform provided realistic-looking characters and background in a familiar setting.

An Irish language storyline was written for each of the platforms and *ABAIR* synthetic speech added. By testing these separately one could adduce, to some extent, the relative merits seen by the learners in each of the features.

Neither of the platforms in their current implementation, however, necessitated synthetic speech for their operationalisation. The architecture of The Language Trap, on which *Digichaint* was based, was such that it was limited to a predefined speech corpus to which the learner could not add new or unpredictable utterances. The excerpt from the *Metropolis Project*, on which *Fáilte go TCD* was based, had no speech accompaniment. In the case of the adaptations, *Digichaint* and *Fáilte go TCD*, natural speech could (in principle) have been recorded and added rather than using synthetic speech. Of course, ultimately, it is envisaged that this kind of platform can be used as a shell which teachers and programme developers can eventually use to build their own language learning materials, adapted to the particular language learning group and its needs. In that context, the use of synthetic speech would be essential. In order to produce a platform for which synthetic speech would be essential for present purposes, however, a third platform ***"Taidhgín"*** was developed.

Taidhgín is a chatbot, as described in Chapter 3.2.3. This dialogue partner was developed using the AIML (Artificial Intelligence Markup Language) programming language and is hosted on the open source platform *Pandorabots*. The graphical interface for *Taidhgín* consists of an animated monkey typing into a laptop. The screen also has an input box for the user to type relatively unpredictable utterances into the system. This input is held on the screen for the user to see and the response given by *Taidhgín* is presented below it in both text and speech forms. The user is also given the option of hearing his/her own typed input spoken by either a male or female synthetic voice. The response strings uttered by *Taidhgín* are generated based on the AIML matching algorithm finding the most suitable predefined response to the input given by the user. It can combine units from the input with its own predefined stored lexicon and output novel utterances:

Taidhgín: Cad is ainm duit?

User: Is mise Máire.

Taidhgín: Deas bualadh leat, **a Mháire**.

It would not be possible to use prerecorded natural voices here, for example, since the number of possible names which may be inputted by the user is limitless. Similarly, it would be unrealistic to try to prerecord dates of birth since one would need a recording of every single day, of every month, for each year back to the age of the oldest living person. *Taidhgín*, therefore, demonstrates the necessity for synthetic speech in a dialogue-based CALL platform since input from the user cannot be anticipated.

Like *Fáilte go TCD*, the *Taidhgín* platform does not have game-like features and its main function is to demonstrate the use of synthetic speech in personalised interactive dialogue-based platforms.

The creation and development of these three platforms with their authentic, personalised formats were the backdrops on which the synthetic speech from *ABAIR* could best be demonstrated. They represent real-life situations where the functionality of the synthetic speech can be judged by teachers and by pupils in order to draw some conclusions on the basic research questions of this thesis, i.e. are these kinds of game-like CALL platforms, using Irish synthetic speech, adequate/useful/enjoyable?

6.3 The Synthetic Voices

The development of TTS synthetic voices for Irish at the Phonetics and Speech Lab., School of Linguistic, Speech and Communication Sciences, TCD is still an ongoing

project. There are two main voices currently available to the public, an Ulster female voice and a Connaught male voice. There are two versions of the Connaught male voice available and these were developed using two different methodologies – unit-selection synthesis and the HMM-based speech synthesis system HTS (see Chapter 2.6.3). The HTS synthesis was used for *Taidhgín* while the other two platforms used unit-selection synthesis. The HTS system, though sometimes lacking the clarity of unit-selection synthesis, has the advantage of generating its output at greater speed. This was considered necessary for the chatbot in order to allow it to produce almost immediate responses to the inputted text giving the sense of real-time conversation.

Because of being confined to just one male and one female synthetic voice for a cast that had more than two characters in both *Digichaint* and *Fáilte go TCD*, it was necessary to manipulate the pitch and speed of the voices so as to differentiate between the characters.

6.4 Testing and Analysis

One of the main aims of the present project was to test the adequacy of the TTS synthesis produced by *ABAIR* for use in game-like CALL platforms. It was being tested from a learner/teacher perspective.

A fundamental principle of the testing was that the speech output should be judged in the context of the environment in which it was likely to be used (see Chapter 3.3.2). It was hypothesised that judgments of the synthetic speech were likely to be influenced by one's overall impression of the platforms (Lim & Chai, 2008). The complicating factor here is that one is attempting to test two unknowns simultaneously – the acceptability of both the linguistic (TTS) and non-linguistic (non-TTS) features of the platforms was unknown. If the speech was to get a negative rating then could part of that negativity be ascribed to a negativity towards other aspects of the platforms? In order to try to overcome this problem, questionnaires were devised which focused separately on various aspects of the

platforms, such as graphics, storyline, standard of the Irish being used. Questions were then posed on the intelligibility, quality and attractiveness of the synthetic voices.

The TTS synthesis is best judged using both qualitative and quantitative measures (Liu et al., 2003). For the purposes of the present study, independent evaluations were conducted firstly using 16-17 year old pupils attending second level schools (N = 252) and secondly using a group of 31 teachers who volunteered to participate in the project. This latter group partook in an online transcription exercise where they were asked to transcribe twenty sentences. They were also asked to rate, on a 5-point Likert scale, the ease with which they could perform the tasks. The transcribed sentences were marked for accuracy. This is regarded as the objective assessment of intelligibility the speech. Consideration was given to the idea of testing the platforms in terms of the extent to which pupils learned some new material from them. This would be of a pre-treatment, treatment and post-treatment type approach (Barr et al., 2005). The idea was rejected since one could not readily ascribe any knowledge which pupils were found to have in a post-treatment test to exposure to that particular platform (see Chapter 3.3.1.1 for discussion). Pupils may have input from numerous other sources as well as from past experience. One could not speculate that the pupils did not have some exposure to the language forms at some other time.

The pupils who participated in the testing came from three school types:

(a) English-medium schools in both rural and urban backgrounds
(b) Gaeltacht schools
(c) Gaelscoileanna (Irish-medium schools outside of Gaeltacht areas)

(see Chapter 2.2.5 for an explanation of the differences between these types of schools)

There was a total of 250 (*Digichaint*), 252 (*Fáilte go TCD*) and 228 (*Taidhgín*) participants. A pilot study was conducted prior to the main study in which a further 30 pupils participated. This was used as a 'dry run' both in terms of testing (1) the robustness of the online data gathering system which was developed specifically for this project, and (2) for ambiguity in the questionnaire items. Adjustments were made where necessary. This group was not included in the figures pertaining to the main study.

Data were accrued by way of 5-point Likert scales. In the case of the teacher data, separate scores were arrived at for their performance in the transcription test and for their reported score for the 'ease' with which they could transcribe each sentence.

Scores gathered by way of Likert scales may be analysed using non-parametric analysis techniques. In the first instance, results were reported in terms of descriptive statistics which were given as the percentages of respondents who marked a questionnaire item at a point from 1 to 5 (typically – strongly disagree; disagree; neutral; agree; strongly agree). Since there was quite a spread in the range of responses, further analysis was carried out by way of Kruskal-Wallis tests in order to examine the relationship between 5 key background factors pertaining to the respondents and their post-game assessment of the platforms. These background factors were gender; school type; frequency with which the learners play computer games; ability to understand Irish; and attitude towards synthetic voices.

A further analysis was carried out using Spearman's rho correlation tests in order to establish the relationship between respondents' attitudes to the different aspects of the platforms including the relationship between their assessment of the graphics, storyline, etc. and their assessment of the synthetic speech in terms of its intelligibility, quality and attractiveness. This analysis tested the hypothesis that one experiences the platforms in a holistic way and that one's impression of one aspect of the platform both influences and is influenced by one's impression of all other aspects of the platform.

6.5 Summary of Teachers' Reactions to Synthetic Speech

The group of teachers who participated in this study were mainly under 40 years of age. Only 13% of them had a preference for the Ulster dialect with more than half of them preferring the Connaught dialect. In this group, 23% came from Gaeltacht areas, 46% from Dublin, and the remainder from various counties throughout the country. They had a negative predisposition towards synthetic speech generally and this negativity concurs with other research findings by Cryer and Home (2009) whose respondents also would prefer human voices but would tolerate synthetic voices if they allowed them to extend their range of talking books. The group had a mixed reaction towards the *ABAIR* synthetic speech. Nobody was completely dismissive of it but only 19% were certain of its suitability. As anticipated, the majority would prefer a human voice but were supportive of the concept of synthetic voices. There was no discussion with the group on the situations where it would be impossible to use pre-recorded human voices. It would seem reasonable to surmise from the response table that teachers would find the voices acceptable in situations where it was impracticable to use human voices. They had a very positive reaction to the quality of the speech and thought that this type of learning environment would be attractive to senior cycle pupils in post-primary schools.

When asked to transcribe twenty sentences, many of which were seen as being potentially problematic from the point of view of a technical synthesis-generation process, the modal response score was 5, representing a fully correct transcription. One can assume from this that the speech is almost fully intelligible to those who have a reasonably good command of Irish. It is worth noting that only a minority expressed a preference for the Ulster dialect prior to hearing the *ABAIR* synthetic speech and some respondents could have had problems because of dialectal issues even if natural voices had been used.

The scores on respondents' reports on the 'ease' with which they could transcribe the sentences were considerably below their performance scores. This suggests that the

speech was more difficult to comprehend than natural speech. It was intelligible to the listeners but it took more effort to process it. This would account for the fact that people tend to prefer human voices to synthetic voices generally, since the latter take more effort to comprehend. An 'ideal' score, indicating complete ease of intelligibility, where respondents could write the sentence correctly and do so with complete ease, was taken as 5. An average or mean score was calculated for the group of respondents, which represented the extent to which they deviated from the ideal score in terms of their performance and the reported degree of ease with which the performance was achieved. The Ulster dialect had the greater deviation from the perfect score. This suggests that some respondents had difficulty with the dialect rather than there being a problem with the quality of the synthetic speech as such. This flags an important issue in the context of a general evaluation of the *ABAIR* output. Due weightings will need to be given to the responses of evaluators from different Gaeltacht areas. Those who are completely familiar with the dialect may give a higher rating to the speech than those who are not. The present sample is too small for one to arrive at a definitive conclusion in this regard but it may act as a pointer for future studies. The average deviation from 5 for the Ulster dialect was 1.37 while the corresponding figure for the Connaught dialect was 0.86. Results from this part of the study point to the speech being fit-for-purpose for use in CALL platforms where the use of pre-recorded speech is impracticable.

6.6 Summary of Learners' Responses

Learners came mainly from fifth year classes in a variety of post-primary schools throughout the country, including English-medium schools, Gaelscoileanna and Gaeltacht schools. There was a mixture of boys and girls but girls were in the majority. They reported a low level of engagement with computer games generally. They were generally favourable towards Irish with almost half of them counting Irish amongst their top three subjects. Almost all had taken higher level Irish for their Junior Certificate and most intended taking higher level Irish for their Leaving Certificate.

A detailed analysis of the opinions of the pupils on each of the three platforms is given in Chapter 5.9. The present summary will confine itself to identifying general trends which emerge from the analysis and reviewing their significance.

One can claim a generally positive view of the platforms. One can have some confidence in the integrity of the responses since they were given totally anonymously online and respondents used the full 1 to 5 Likert scale that was available to them.

The graphics got a high rating across all three platforms. It is interesting to note that the off-the-shelf *Taidhgín* got an approximately equal rating to the other two which were highly stylised. The vast majority felt that there were no technical difficulties in manoeuvring through the *Digichaint* game. There was an acceptance of the authenticity of the movements of the avatars and their alignment with the speech. The sense of playfulness of the monkey got the highest rating of any feature of the three platforms. *Taidhgín* got a much higher rating than *Fáilte go TCD* for its attractiveness and respondents would prefer it as a tool on which to practise language exercises. This may be ascribed to the fact that it is more highly interactive, allowing greater participation by the learner, a greater sense of control, and ultimately a stronger impression of being involved in a two-way communicative experience. The vast majority of respondents found that these types of platforms would be '*useful*' or '*very useful*' as an aid to their practising oral or aural skills.

The evaluation of *Digichaint* included additional game-related questionnaire items as it was the only one of the three platforms which had strong gaming features. It was found to be technically very acceptable to respondents, the plot was clear, the hypertext facility was useful, it was easy to manipulate and generally enjoyable. There was agreement that the plot was credible, the game held their attention, they were more focused on the plot of the game than they were on the language being used and there was a good balance between enjoyment and language learning.

The three platforms conformed to the general criteria, derived from the review of literature, which are regarded as most pertinent to the development of stand-alone CALL platforms (see Chapter 3.3.1.2). These then can be regarded as suitable backdrops on which to display and evaluate the fitness-for-purpose of the TTS synthesis.

The difficulty level of the language used was about right and those who were not satisfied with it were about equally divided as to whether they thought it was too difficult or too easy.

There was divided opinion as to whether the speech was sufficiently clear so as to be intelligible. Almost 10% '*completely agreed*' that it was intelligible while 6% '*completely disagreed*'. The majority, however, were on the positive side of neutral. Similarly, there was a complete range of opinions as to whether the dialects posed particular problems or as to whether it would be easier to understand natural voices. The quality and attractiveness of the synthetic speech got an equally mixed reaction. It is interesting to note that the highest rating for intelligibility was given to *Taidhgín*. There may have been a number of contributory factors to this rating. The *Taidhgín* voice was generated using HTS synthesis and there was no post production manipulation of pitch and speed. Learners had control over their own input and therefore were more likely to focus attention on the response. When the topic is determined by the learner, then he/she can focus on the most communicatively salient points in a personalised response. This comes close to what has been referred to as "negotiation of meaning" (Little, 1994) and is most likely to be associated with the high intelligibility rating given to the speech of *Taidhgín*. It is also worth noting that the text was printed on the screen in both the *Digichaint* and *Taidhgín* platforms but not for *Fáilte go TCD*. This may have added to the ratings of intelligibility of the speech in the former two platforms.

In the light of the synthetic speech being highly intelligible to teachers and to some pupils it was necessary to further analyse these results in order to examine the respondents'

background characteristics which were associated with acceptance or rejection of the synthetic voices.

6.7 Learners' Backgrounds and their Responses

One of the dangers of interpreting responses gathered through Likert type questionnaires is that respondents may give the 'polite' answer or the answer which they see as being expected of them (Worcester & Burns, 1975). Alternatively they may tend towards the cautious answer and adopt neutral or midpoint on the scale (DeVellis, 2012; Krosnick & Presser, 2010). In this case, there was a complete spread of opinions, which suggests that responses were not constrained by the desire to be polite or to be cautious.

In order to come to a better understanding of this diversity, the relationship between respondents' background characteristics and response patterns was tested for statistical significance using the Kruskal-Wallis test.

While more girls than boys participated in the study, there were almost no gender differences in the responses. Girls gave a higher ranking to the graphics in *Digichaint* but, overall, one could not claim that there are major gender differences in respondents' reactions to the three platforms.

The key underlying factor in determining their responses to all three platforms was their prior attitude towards synthetic voices generally. Those who had a negative attitude towards synthetic voices generally were more likely to be negative in their attitude towards many features of the platforms, including the voices. This suggests that there may be a low level of acceptance, amongst some respondents, of using digital applications as a learning medium. A significant percentage (73%) reported that they '*seldom*' or '*never*' played computer games. Computers in classrooms are a relatively

recent phenomenon and it is still rare for pupils to have personal computers in class. It may be that they are not regarded by some as having much relevance to learning. These people may not be "comfortable" with the use of certain technologies and may confine their use to little more than social media.

Related to the question of predisposition towards synthetic voices was the question of frequency with which pupils played computer games. Those who never played games showed least regard for the *Digichaint* game. This may be associated with a lack of skill in manoeuvering through the game or a lack of enthusiasm for any kind of computer game. Those who played games infrequently gave the highest rating to many of the games' features – a much higher rating than that given by regular or frequent players. This suggests that those who play games infrequently have the basic computer skills but are not comparing *Digichaint* with commercial computer games. This opinion is supported by the finding that frequency of playing games was of lesser importance in respondents' reaction to *Fáilte go TCD*, which requires relatively little computer manipulation by the learner.

School type and level of understanding of Irish may be jointly indicative of respondents' level of proficiency in Irish. Those in Gaeltacht schools and in Gaelscoileanna reported, as expected, a higher level of competence in Irish. The relationship between this factor and their responses to the various aspects of the platforms was complex. There was a greater enthusiasm for the platforms and for their future potential amongst those with lower levels of Irish competence. It would appear that those at the highest levels of competence saw little use for the platforms presented to them for their own purposes. Those from the highest competency level had no difficulty with the voices or the dialects. These results demonstrate the importance of relevance and engagement or "flow" (Csíkszentmihályi, 1988) if learners are to become engaged with a CALL platform. Those who saw little relevance for the present platforms gave lower rankings to their various features. The platforms represented an appropriate level of challenge for those in English-medium schools who had an intermediate level of competence in Irish. They

consequently had a higher level of engagement with the platforms and showed greater appreciation of their graphics, their hypertext features, etc.

The extent to which one becomes involved with the platforms colours one's attitude towards almost every aspect of that platform. Results from the Spearman's rho correlation tests (which investigated the mutual influences amongst pupils' responses to each of the post-game questionnaire items) show a significant relationship between respondents' opinions of all the features of the games. If one gave a high ranking to one item then they tended to give a high ranking to all of the other items and vice versa. This result gives strong support to Lim and Chai's (2008) proposal that playing computer games is a holistic experience. Careful consideration, therefore, needs to be given to the design and relevance of every aspect of the CALL platform. A further discussion follows on platform design in Section 6.9, which makes proposals for the development of such a platform.

The results of the analysis of both teacher and pupil generated data point to the adequacy of the *ABAIR* system for the type of interactive CALL platforms developed for the present project.

This result has very significant implications for the development of CALL materials for Irish. It suggests that learner-computer interactions could reach a new level from that which is possible if only pre-recorded set texts are available. Synthetic speech allows for the possibility of an infinite number of novel utterances which can give unique direction to the learner-computer interactions. Detailed methodological considerations of the construction of these dialogue patterns are beyond the scope of the present thesis. The *Taidhgín* platform gave a basic prototype example of the possibilities within a restricted domain. These domains may be extended and the architecture for dialogue construction expanded and advanced.

The platforms developed for the present project were based on the assumption that their primary educational goal was the development of aural and oral skills and that such platforms would become another source of comprehensible input and be complementary to current classroom practices. An integrated synthetic speech system offers new possibilities for personalised language learning opportunities when combined with Artificial Intelligence (AI) systems. Some consideration is given to the development of such a system in Section 6.9 below.

6.8 Summary of Hypotheses Findings

The following is a brief summary of the results of testing the hypotheses. These hypotheses were set out first in Chapter 3.4. All hypotheses, which were derived from the review of the literature, were largely supported by the data collected for this study.

Hypothesis 1

Digichaint, as an interactive game-like platform, will conform to the criteria for effective CALL platforms (as set out in Table 3.2).

Hypothesis 1 was laid out in terms of nine subsections. Each subsection was tested based on data from pupils' *Digichaint* evaluation. A positive support for each subsection of Hypothesis 1 was found (see Chapter 5.9.4.1).

Hypothesis 2

The synthesised speech is readily intelligible for respondents with a high Irish language comprehension ability.

Hypothesis 2 was strongly supported by reference to results from the teachers' transcription exercise. The teachers transcribed 20 sentences with near perfect results (see Chapter 4.9.3.1).

In relation to the learners' responses, I attempted to glean support for the hypothesis by anticipating between-school differences and between-self-assessed language ability differences as well as a number of post-game questionnaire items which specifically referred to the TTS synthetic voices. There were almost no differences between the learners' ratings of speech intelligibility, regardless of the type of school they came from, or of how they had ranked their language ability on the Likert scale. It is noted, however, that most of the learners had a relatively high language ability level so that, on reflection, it is not surprising that few statistically significant differences existed between the various groups (see Chapter 5.9.4.2 for discussion). This must be seen, however, in the context of the descriptive statistics which found that 58% (average across the three platforms) found the synthetic speech sufficiently clear to make the speech intelligible.

Hypothesis 3

The synthesised voices are regarded by a majority of respondents as having a high quality and attractiveness.

Teachers were largely positive about the 'quality' of the voices but they did not regard them as 'attractive' and would have preferred human voices. There was no discussion, however, on the necessity to use synthetic voices in areas where the text would not be predictable. Teachers were very positive about the value of an interactive language learning platform and since they regarded the voices as being of high quality one could argue that these voices would be acceptable to teachers where it is not possible to use pre-recorded human voices (see Chapter 4.10).

The pupils' responses, as shown in Chapter 5.9.4.3, largely mirrored the responses of the teachers. They regarded the synthetic voices as having 'high quality' but they would prefer human voices in terms of their 'attractiveness'. They were not dismissive of the use of synthetic voices and again, one could glean that such voices would be acceptable where it is not possible to use human voices.

Hypothesis 4

A majority of respondents regard the synthesised voices as being acceptable in the context of the CALL platforms in which they are presented.

Support for this hypothesis can be taken from the fact that 71% of teachers gave a rating of 4 or 5 on the Likert scale to the '*quality*' of the synthetic voices (Table 4.7). While one notes teachers' preference for human voices, it is also important to note that 51% of teachers gave a rating of 4 or 5 on the Likert scale to the idea that they considered that the voices would be '*attractive*' for pupils (see discussion in Chapter 4.10).

A majority of pupils '*agreed*' or '*strongly agreed*' that the voices were acceptable in the context of a CALL platform. It is also noted that a majority of pupils regarded the voices as being '*adequate*' or '*totally adequate*' for a game-like platform (see Chapter 5.9.4.4).

Hypothesis 5

There is a statistically significant relationship between the measured reactions of the respondents to non-linguistic elements of the platforms (such as plot, ease of manoeuvring characters, legitimacy of narrative and standard of Irish language used) and their opinions on the quality, intelligibility and attractiveness of the synthesised voices used in each platform.

This hypothesis was tested in Chapter 5.9.3.3.6 by way of Spearman's Rank Correlation Coefficient. An inspection of the tables of correlation coefficients shows that there was a significant statistical relationship between almost all items in each of the three platforms (see Appendix G). This shows that one's opinion on the intelligibility, quality and attractiveness of the synthetic voices is related to one's opinion of the platform or context in which the voices were used and suggests that one's evaluation of the synthetic voices is context-dependent. It highlights the need for high quality platforms if synthetic speech is to be integrated into a CALL platform.

Hypothesis 6

There is a significant relationship between respondents' predisposition towards synthesised voices in general and their attitude towards the Irish synthesised voices being tested.

The data for learners' opinions of the synthetic voices, as used in each of the 3 platforms, give strong support to this hypothesis. In the case of *Fáilte go TCD* one's predisposition towards synthetic voices had a highly significant effect on all aspects of the synthetic voices which were enquired into (see Chapter 5.9.4.6). There was a similar result for the *Taidhgín* platform, although one of the items (Item 11_2) fell just short of statistical significance at $p<0.05$. The outcome for the *Digichaint* platform, however, was less decisive, with a significant relationship being shown for three of the six items being enquired into. It is clear, however, that one's opinion on the intelligibility, quality and attractiveness of the voices was significantly related to one's predisposition towards synthetic voices.

6.9 Towards a Framework for the Development of a CALL Environment

This thesis has concerned itself with the development of stand-alone CALL platforms which take on the characteristics of what has been dubbed 'serious games' (Susi et al., 2007). These form but a very small part of the entire plethora of computer mediated activities which come under the heading of CALL. Developments in natural language processing and in artificial intelligence have opened a whole new range of possibilities for CALL.

The proposal being forwarded here comes with a caveat. No CALL program will be a panacea for language teaching/learning. To overstate the value of CALL would be to do it a disservice by building false hopes leading to a rejection of the entire concept by many teachers and learners. Not all learners are entirely enamoured with computer-assisted learning. Jones and Cross (2009) suggested that new technologies such as blogs, wikis and virtual worlds were less used by students of third level institutions than might have been expected. A huge divergence was found among students in their attitudes towards computer usage with some being quite technophobic. Kennedy et al. (2007) similarly found that Australian third level students were less interested in Web 2.0 technologies than teachers may have imagined. O'Donnell and Sharp (2012) found little difficulty in arguing for the continuance of the employment of lecturers on the basis that technology is seen as an enhancement of their learning experiences and not a replacement for face-to-face contact with lecturers and peers. No surveys have been conducted on the use of learner controlled e-learning in post-primary schools in Ireland but anecdotal evidence would suggest that it is very small. It is in this context, and in the context of the theoretical considerations discussed in this thesis, that the following proposals for the development of a CALL environment are being proffered.

Before proceeding any further it should be mentioned that there tends to be a problem with precise definitions of CALL terminology, such as 'platforms' and 'environments'.

Leakey (2011) refers to platforms as the skeleton and the software as the muscle. In this thesis a platform is being referred to as any functional piece of software which has been specifically designed to facilitate language acquisition. Platforms may be stand-alone or may interconnected to each other in the sense of sharing databases or software features.

The key elements that are proposed as necessary in the construction of an ideal CALL environment are set out in diagrammatic format in Figure 6.1.

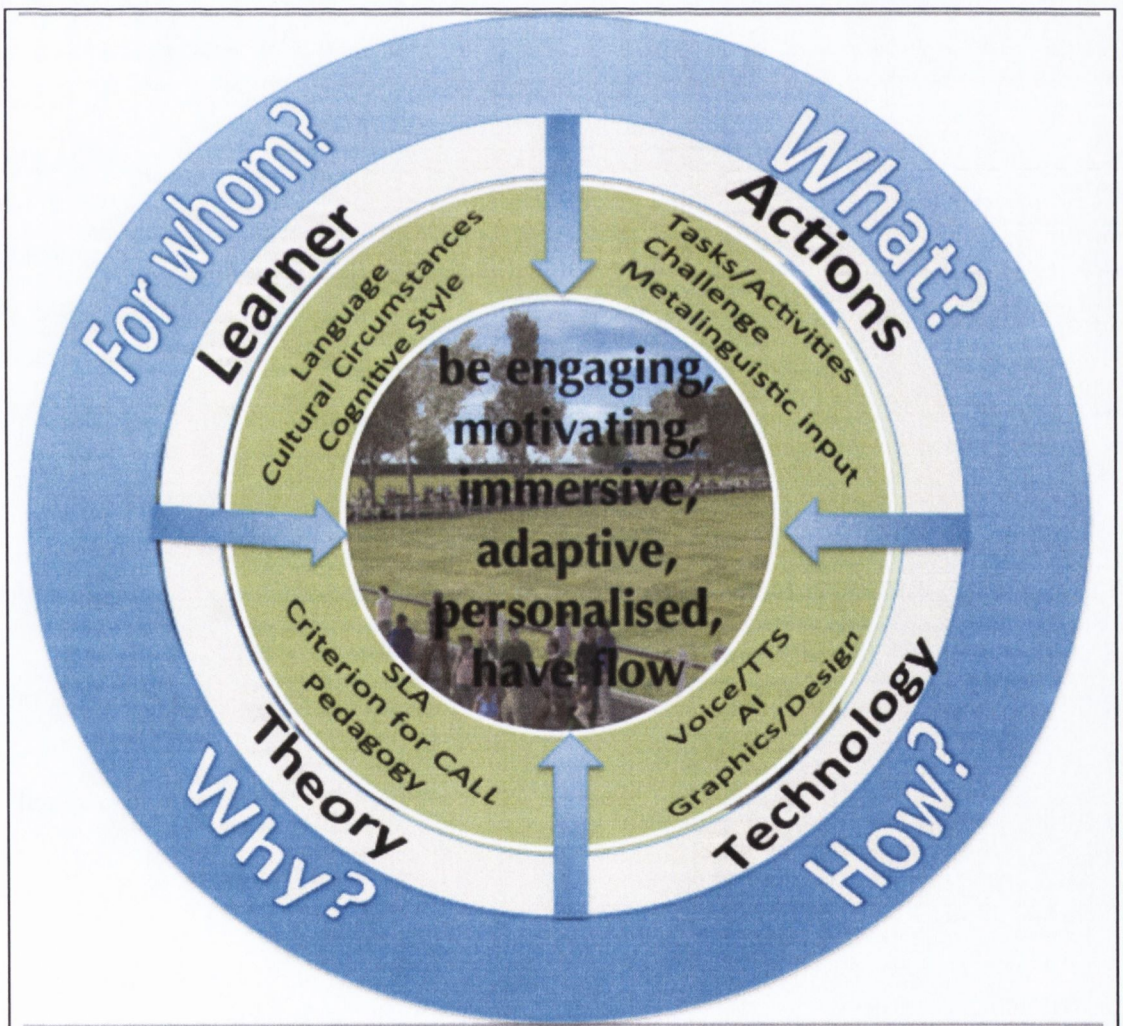


Figure 6.1. Blueprint for Future CALL Environment.

Figure 6.1 is being referred to as a CALL environment and represents the entire set of interconnections which go into determining the list of activities which form a coherent CALL whole. These activities should share to the greatest extent possible the following features: be engaging, motivating, immersive, adaptive, personalised, and have flow. Individual activities, be those games, life-simulating activities or other, are being referred to as platforms. A platform may be as simple as a talking dictionary or as complex as a 3D virtual problem-solving activity.

The shape of the CALL environment is determined by four key considerations, each belonging to disciplines which are distinct from one other. Each of these areas of consideration is referred to here as a module. A module may be thought of in terms of a distinct body (of knowledge) which circulates around a centre and which reacts with other circulating modules and brings a unique contribution to the environment in which it operates. Each of the modules needs its own particular practitioners as they represent areas of research which are completely distinct from one other. Each module may inform and be informed by the developments in each of the other modules. These four modules are:

1. The *Why* module, or the theoretical underpinning for the entire project
2. The *What* module, refers to the activities which one wishes to promote in the CALL environment
3. The *How* module refers to the technological inputs which are possible. It may bring the desirable activities into fruition
4. The *For Whom* module refers to the learner in his/her unique psycho-social environment

The ***Why*** module is the raison d'être for the entire enterprise and draws from research in areas such as Second/Foreign Language Acquisition, Pedagogy, Criteria for CALL, etc. Tam and Taher (2012) see theory in terms of clarifying "your underlying belief in how

knowledge is acquired, how language is learnt and how computer skill is mastered" (p. 132). In the present thesis a task-based communicative approach to language acquisition was taken.

The theory is localised to a specific situation by the ***For Whom*** module, described as "Learner" for short. This takes account of factors such as the cultural circumstances in which the platform is being constructed. It also takes account of individual differences. Brantmeier, Flores and Romero-Ghiretti (2006) list beliefs, aptitudes, learning strategies, attitudes, motivation, age, personality, gender and learning styles as being among the variables which are vital to the development of IT projects. Fatemi Jahromi and Salimi (2013) also refer to the human factors as having a predictive and determining effect on the acceptance or rejection of technology. Platforms need to be personalised to the needs and characteristics of the target groups of learners.

The ***What*** module, described as "Actions", refers to the specific games or activities which the pedagogues deem desirable. It refers to the most desirable activities which should be hosted in the particular environment being developed. Here we ask questions such as: what activities should we aim to develop for a particular set of learners on whom we are focusing. How do we get an authentic narrative? How much emphasis should be put on developing metalinguistic awareness? Should we have some kind of feedback for the learners, as suggested by Livingstone (2012)?

The needs of the *What* module can only be met by what is made possible in the ***How*** module. This module represents the technical computer-based disciplines such as Artificial Intelligence, Graphics, Voice Research, Computer-based Design, Natural Language Processing, etc. It is referred to here as the "*How*?" or the technical expertise required to implement the specific activities, for specific groups, in line with accepted pedagogical theory. Certain desirable activities may not be technically feasible. Technological progress is rapid. In the context of the present thesis focus was put on

synthetic speech and its contribution to what activities are now made possible because of the synthetic speech having reached a point in its development where it is deemed 'fit-for-purpose' in particular game-like activities. In future there may be developments in, say, speech recognition technology which will influence the *What* module, i.e. an increased range of activities may be made possible if it is of a high enough quality to allow a relevant platform to be developed successfully. Artificial Intelligence can be used to develop materials which have a much higher degree of interactivity between user and computer.

In the earlier sections of this thesis the question was posed as to whether the pedagogue or the engineer led developments in CALL (see Chapter 2.3.2). There was a suggestion that CALL developed from pedagogies latching on to technological developments which were not specifically intended to have any particular relevance to language teaching. The present model describes a symbiotic relationship between the four modules which go to make up the CALL environment. There is a two-way relationship between each of the modules. Their interactions may be complex and will vary from one socio-cultural/economic situation to another. If the CALL environment can be visualised as the centre of Figure 6.1 then each of the modules circumnavigates this central environment in which the platforms are hosted. Each shines its own distinctive form of radiance on the whole.

The model presented here is dynamic and represents an ever-evolving unified CALL environment in which several interdependent platforms may be hosted. It is an idealised blueprint, which may act as a unifying framework for various CALL activities. It is neither context- nor language-specific. Within the *For Whom* module allowance is made for variation in learners and this can also account for variations in the language being promoted. What is essential is that each activity must comply with a common architecture in order to ensure technological compatibility. The architecture for this environment can be sufficiently flexible to allow for language variation and for other learner-specific variations. Resources developed for one platform must have the capacity to be shared or

incorporated into parallel platforms. For example, vocabulary items or turn-taking patterns for a discourse developed for one platform must have direct transferability to another platform being developed within the same architectural framework. The blueprint can develop from very small beginnings. The resources required for a large scale development may not be available, yet some relatively easily produced materials may be possible, depending on the available technical support and resources. These may subsequently be expanded, with databases etc. shared between all contributors provided they comply with the technical compatibility requirement.

An example of the kind of development being referred to here is the *Taidhgín* chatbot being upgraded using AI capabilities to include a greater domain expansion, feedback of a corrective or motivational nature such as those developed by Ward (2007), and learners' progress monitoring (Livingstone, 2012). It is envisaged that the database developed could be extended to the *Digichaint* type setting where, because of the availability of synthetic speech, a new and expanded level of personalisation is possible. Learners would no longer be confined to predetermined utterances but could now have their own unique conversational input in playing the game. The *Digichaint* platform could, in turn, be transposed into a SecondLife-type environment as in the environment of *Fáilte go TCD*, or the *Metropolis* project. The relatively large database accumulated from the previous two stages is immediately available to the new environment. Each stage has a legitimacy in itself but can be developed into a further stage without reinventing most of its key components. The platform would be in a state of continuous evolution depending on feedback from continuous evaluations.

The process model represented in Figure 6.1 is akin to the processes of "internationalisation" and "localisation" – processes which, when combined, are being referred to by IBM and Sun Microsystems as "globalisation". The Localization Industry Standards Association (LISA) defines "internationalisation" as "the process of generalising a product so that it can handle multiple languages and cultural conventions

without the need for re-design. Internationalization takes place at the level of program design and document development".

The model of CALL development being proposed here is sufficiently general so that it can provide the framework on which localised platforms can be developed. LISA defines localisation as "taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold". Chandler and O'Malley Deming (2011) refer to the opportunities for the internationalisation and localisation of game software. Just as CALL has associations with game development, it is now possible that CALL materials may be developed which have international possibilities. Such possibilities would bring the development of high quality CALL materials to a point where they may become economically viable.

The TTS synthetic voices as developed by *ABAIR* have already reached the point where they can make a very significant contribution to the development of highly interactive CALL materials. The development of such materials needs to be done in a systematic way rather than by piece-meal individual projects which are unable to interconnect, to share databases or to cooperate with one another. Planners of the development of this environment must include those with expertise in each of the four areas, or modules, identified above. Ideally such development should be done in cooperation with international expertise so that factors such as graphics development, dialogue structure and artificial intelligence resources can be shared.

References

Abu Shawar, B., & Atwell, E. (2007a). Chatbots: are they really useful? *LDV-Forum: Zeitschrift Für Computerlinguistik Und Sprachtechnologie*, *22*(1), 29–49.

Abu Shawar, B., & Atwell, E. (2007b). Fostering language learner autonomy via adaptive conversation tutors. In *Proceedings of Corpus Linguistic (CL'07)*. Birmingham, UK.

Agard, F. B., Clements, R. J., Hendrix, W. S., Hocking, E., Pitcher, S. L., van Eerden, A. V., & Doyle, H. G. (1945). A survey of language classes in the army specialized training program. *The Modern Language Journal*, *29*, 155–160. doi:10.1111/j.1540-4781.1945.tb03356.x

Al-Hejin, B. (2005). Attention and awareness: Evidence from cognitive and second language acquisition research. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *4*(1), 1–22.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202–232. Retrieved from http://llt.msu.edu

Al-Wabil, A., Al-Khalifa, H., & Al-Saleh, W. (2007). Arabic text-to-speech synthesis: A preliminary evaluation. In C. Montgomerie & J. Seale (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 4423–4430). Chesapeake, VA.

Alvarez, Y. V., & Huckvale, M. (2002). The reliability of the ITU-T P. 85 standard for the evaluation of text-to-speech systems. In *Proceedings of International Conference on Speech and Language Processing (ICSLP-2002)*. Denver, Colorado.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Amory, A., Naicker, K., Vincent, J., & Adams, C. (1999). The use of computer games as an educational tool: Identification of appropriate game types and game elements. *British Journal of Educational Technology*, *30*, 311–321. doi:10.1111/1467-8535.00121

An Coimisinéir Teanga. (2008). General questions about the official languages act. *An Coimisinéir Teanga: Protecting Language Rights*. Retrieved October 30, 2013, from http://www.coimisineir.ie/downloads/FAQ001GeneralQuestionsAboutTheOfficialLanguagesAct.pdf

Andersson, S., Yamagishi, J., & Clark, R. A. J. (2012). Synthesis and evaluation of conversational characteristics in HMM-based speech synthesis. *Speech Communication*, *54*, 175–188. doi:10.1016/j.specom.2011.08.001

Andraszewicz, S., Yamagishi, J., & King, S. (2011). Vocal attractiveness of statistical speech synthesisers. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)* (pp. 5368–5371). Prague, Czech Republic. doi:10.1109/ICASSP.2011.5947571

Antons, J.-N., Laghari, K. ur R., Arndt, S., Schleicher, R., Möller, S., O'Shaughnessy, D., & Falk, T. H. (2013). Cognitive, affective, and experience correlates of speech quality perception in complex listening conditions. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)* (pp. 3672–3676). Vancouver, BC: IEEE. doi:10.1109/ICASSP.2013.6638343

Atkinson, R. C. (1974). Teaching children to read using a computer. *American Psychologist*, *29*, 169–178. doi:10.1037/h0036071

Auclerc, N., Bramoullé, A., Favereau, F., Gourmelon, H., Lavanant, P., Louis, F., … Siroux, J. (2000). Bilingual spoken dictionary based on speech synthesis in Breton. In *Proceedings of the Sixth Annual Conference of the North American Association for Celtic Language Teachers* (pp. 11–17). Limerick, Ireland: North American Association for Celtic Language Teachers.

Auld, G. (2002). The role of the computer in learning Ndj Bbana. *Language Learning & Technology*, *6*(2), 41–58. Retrieved from http://llt.msu.edu/vol6num2/auld/

Bailin, A. (1995). AI and language learning: Theory and evaluation. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology* (pp. 327–343). Mahwah, NJ: Lawrence Erlbaum Associates.

Barab, S. A., Arici, A., & Jackson, C. (2005). Eat your vegetables and do your homework: A design-based investigation of enjoyment and meaning in learning. *Educational Technology: The Magazine for Managers of Change in Education*, (1), 15–20.

Barr, D., Leakey, J., & Ranchoux, A. (2005). Told like it is! An evaluation of an integrated oral development pilot project. *Language Learning & Technology*, *9*(3), 55–78. Retrieved from http://llt.msu.edu

Bax, S. (2003). CALL-past, present and future. *System*, *31*, 13–28. doi:10.1016/S0346-251X(02)00071-4

Beatty, K. (2003). *Teaching and researching computer-assisted language learning*. London: Pearson Education.

Beatty, K. (2013). *Teaching & researching: Computer-assisted language learning* (2nd ed.). New York: Routledge.

Bergström, H. (2007). *Evaluation of a computer-assisted language learning system for Swedish language learners* (Master's thesis, KTH CSC, Royal Institute of Technology, Stockholm). Retrieved from http://www.speech.kth.se/prod/publications/files/3041.pdf

Black, A. W., & Campbell, N. (1995). Optimising selection of units from speech databases for concatenative synthesis. In *Proceedings of Eurospeech95* (pp. 581–584). Madrid, Spain.

Black, A. W., Zen, H., & Tokuda, K. (2007). Statistical parametric speech synthesis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)* (pp. IV–1229–IV–1232). Honolulu, HI: IEEE. doi:10.1109/ICASSP.2007.367298

Blake, R. J. (2007). New trends in using technology in the language curriculum. *Annual Review of Applied Linguistics*, *27*, 76–97. doi:10.1017/S0267190508070049

Brantmeier, C., Flores, L., & Romero-Ghiretti, G. (2006). Theory driven technologies: Frameworks for individual language learners. *The Reading Matrix: An International Online Journal*, *6*(3), 299–309.

Breen, M. P., & Candlin, C. N. (1987). Which materials? A consumer's and designer's guide. In L. E. Sheldon (Ed.), *ELT Textbooks and materials: Problems in Evaluation and Development* (pp. 13–28). London: Modern English Publications.

Breuer, J., & Bente, G. (2010). Why so serious ? On the relation of serious games and learning. *Eludamos. Journal for Computer Game Culture*, *4*(1), 7–24. Retrieved from http://www.eludamos.org

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). White Plains, NY: Pearson Education.

Bunreacht na hÉireann. Constitution of Ireland, enacted in 1937 (1937). Retrieved from http://www.taoiseach.gov.ie/eng/Historical_Information/The_Constitution/

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. Harlow, UK: Longman.

Campbell, N. (2007). Evaluation of speech synthesis. In L. Dybkjær, H. Hemsen, & W. Minker (Eds.), *Evaluation of text and speech systems* (pp. 29–64). The Netherlands: Springer.

CCIR-5. (1999). *User attitudes towards real and synthetic speech*. University of Edinburgh. Retrieved from http://spotlight.ccir.ed.ac.uk/public_documents/technology_reports/No.11 TTS.pdf

Chandler, H. M., & O'Malley Deming, S. (2011). *The game localization handbook* (2nd ed.). Sudbury, MA: Jones & Bartlett Learning.

Chang, C.-K., & Hsu, C.-K. (2011). A mobile-assisted synchronously collaborative translation–annotation system for English as a foreign language (EFL) reading comprehension. *Computer Assisted Language Learning*, *24*, 155–180. doi:10.1080/09588221.2010.536952

Chapelle, C. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2*(1), 22–34.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge: Cambridge University Press.

Chapelle, C., & Jamieson, J. (2008). *Tips for teaching with CALL: practical approaches to computer-assisted language learning*. New York: Pearson Longman.

Chen, H. H.-J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL*, *23*, 59–78. doi:10.1017/S0958344010000285

CILAR. (1973). *Report of committee on Irish language attitudes research*. Dublin: Government of Ireland.

Clark, R. A. J., & Dusterhoff, K. E. (1999). Objective methods for evaluating synthetic intonation. In *Proceedings of Eurospeech 1999, Volume 4* (pp. 1623–1626). Budapest, Hungary.

Clark, R. A. J., Podsiadlo, M., Fraser, M., Mayo, C., & King, S. (2007). Statistical analysis of the Blizzard Challenge 2007 listening test results. In *Proceedings of the Sixth ISCA Workshop on Speech Synthesis*. Bonn, Germany.

Clark, R. A. J., Richmond, K., & King, S. (2004). Festival 2-build your own general purpose unit selection speech synthesiser. In *Proceedings of the Fifth ISCA Workshop on Speech Synthesis* (pp. 173–178). Pittsburgh, USA.

Clark, R. C., & Mayer, R. E. (2008). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (2nd ed.). San Francisco: Pfeiffer.

CLEF. (1997). CLEF: French grammar package. *Camsoft*. Retrieved October 20, 2013, from www.camsoftpartners.co.uk/clef.htm

COGG. (2013). An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta. Retrieved October 30, 2013, from www.cogg.ie

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: John Wiley & Sons.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2011). Strategies and instruments for data collection and researching. In *Research methods in education* (7th ed.). New York: Routledge.

Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal, 23*, 477–497. doi:10.11139/cj.23.3.477-497

Comas-Quinn, A., Mardomingo, R., & Valentine, C. (2009). Mobile blogs in language learning: making the most of informal and situated learning opportunities. *ReCALL, 21*, 96–112. doi:10.1017/S0958344009000032

Coniam, D. (2008). An evaluation of chatbots as software aids to learning English as a Second Language. *The Eurocall Review*, (13). Retrieved from http://www.eurocall-languages.org/review/13/index.html#coniam

Conlan, O., O'Keeffe, I., Brady, A., & Wade, V. (2007). Principles for designing activity-based personalized e-learning. In *Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (pp. 642–644). Niigata, Japan: IEEE. doi:10.1109/ICALT.2007.209

Cooke-Plagwitz, J. (2008). New directions in CALL: An objective introduction to second life. *CALICO Journal, 25*, 547–557. doi:10.11139/cj.25.3.547-557

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 849–874). Amsterdam: Elsevier.

Crabtree, M., Mirenda, P., & Beukelman, D. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, *6*, 256–261. doi:10.1080/07434619012331275544

Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage.

Cryer, H., & Home, S. (2008). *Exploring the use of synthetic speech by blind and partially sighted people*. RNIB Centre for Accessible Information, Birmingham: Literature review #2.

Cryer, H., & Home, S. (2009). *User attitudes towards synthetic speech for talking books*. RNIB Centre for Accessible Information, Birmingham: Research report #7.

Cryer, H., & Home, S. (2010). *Review of methods for evaluating synthetic speech*. RNIB Centre for Accessible Information, Birmingham: Technical report #8.

Csíkszentmihályi, M. (1988). The flow experience and its significance for human psychology. In M. Csíkszentmihalyi & I. S. Csíkszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 15–35). Cambridge: Cambridge University Press.

CSO. (2011). Daonáireamh na hÉireann: Cainteoirí Gaeilge. *Central Statistics Office*. Retrieved February 25, 2013, from http://www.cso.ie/en/media/csoie/census/documents/census2011profile9/Profile 9 Irish speakers - Combined document.pdf

CSO. (2012). *Profile 9: What we know*. Dublin. Retrieved from http://www.cso.ie/en/media/csoie/census/documents/census2011profile9/Profile,9,What,we,know,Commentary,for,web.pdf

Dagger, D., Wade, V., & Conlan, O. (2005). Personalisation for all: Making adaptive course composition easy. *Journal of Educational Technology & Society*, *8*(3), 9–25.

Dalton, M., & Ní Chasaide, A. (2007). Melodic alignment and micro-dialect variation in Connaught Irish. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: Studies in word and sentence prosody* (pp. 293–315). Berlin: Mouton de Gruyter.

Danan, M. (2004). Captioning and subtitling: undervalued language learning strategies. *Meta: Translators' Journal, 49*, 67–77. doi:10.7202/009021ar

Davies, G. (2006). Computer-assisted language education. In *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. doi:10.1016/b0-08-044854-2/00973-1

Davies, G., & Bruzzone, M. (2004). Fun with Texts. *Maidenhead: Camsoft*. Retrieved October 30, 2013, from http://www.camsoftpartners.co.uk/fwt.htm

Davies, G., Walker, R., Rendall, H., & Hewer, S. (2012). Introduction to Computer Assisted Language Learning (CALL). Module 1.4 in Davies G. (Ed.), *Information and Communications Technology for Language Teachers (ICT4LT)*. Slough, Thames Valley University [Online]. Retrieved from http://www.ict4lt.org/en/en_mod1-4.htm

deHann, J., Reed, M. W., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning and Technology, 14*(2), 74–94. Retrieved from http://llt.msu.edu/

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell Publishing.

Delogu, C., Conte, S., & Sementina, C. (1998). Cognitive factors in the evaluation of synthetic speech. *Speech Communication, 24*, 153–168. doi:10.1016/S0167-6393(98)00009-0

Department of Education and Skills. (1995). *Circular letter M10/94*. Athlone, Co. Westmeath. Retrieved from http://www.education.ie/en/Circulars-and-Forms/Active-Circulars/ppc10_94.pdf

Department of Education and Skills. (2013). Data on individual schools. Retrieved October 30, 2013, from http://www.education.ie/en/Publications/Statistics/Data-on-Individual-Schools

Derryberry, A. (2007). *Serious games: Online games for learning*. Retrieved from http://www.adobe.com/resources/elearning/pdfs/serious_games_wp.pdf

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). California: Sage Publications.

Devitt, S. (1997). Interacting with authentic texts: Multilayered processes. *The Modern Language Journal, 81*, 457–469. doi:10.1111/j.1540-4781.1997.tb05512.x

Dewey, J. (1910). *How we think*. Boston: D.C.Heath & Company.

Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role playing games (MMORPGs) foster intrinsic motivation. *Educational Technology, Research, and Development*, *55*, 253–273. doi:10.1007/s11423-006-9004-7

Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, *42*, 456–469. doi:10.1111/j.1467-8535.2009.01032.x

Dickinson, L. (1995). Autonomy and motivation: A literature review. *System*, *23*, 165–174. doi:10.1016/0346-251X(95)00005-5

Disseldorp, B., & Chambers, D. (2002). Independent access: Which students might benefit from a talking computer? In S. McNamara & E. Stacey (Eds.), *Untangling the Web-Establishing Learning Links: Proceedings of the ASET International Conference*. Melbourne, Australia.

Dörnyei, Z. (2003). *Questionnaires in second language research: Constuction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.

Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology*, *7*(3), 50–80. Retrieved from www.llt.msu.edu

Drager, K. D. R., & Reichle, J. E. (2001). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*, *17*, 109–118. doi:10.1080/aac.17.2.109.119

Drezek, J. (2007). *Adult ESOL reading comprehension and text-to-speech software* (Master's thesis, The University of Texas at Arlington). Retrieved from https://dspace.uta.edu/

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning and Technology*, *13*(3), 66–86. Retrieved from www.llt.msu.edu

Edwards, V., Pemberton, L., Knight, J., & Monaghan, F. (2002). Fabula: A bilingual multimedia authoring environment for children exploring minority languages. *Language Learning & Technology*, *6*(2), 59–69. Retrieved from http://llt.msu.edu/vol6num2/edwards/

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, *2*(1), 45–60. Retrieved from http://llt.msu.edu

Elkind, J. (1998). Computer reading machines for poor readers. *Perspectives*, *24*(2), 9–14.

Elkind, J., Cohen, K., & Murray, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia*, *43*, 238–259. doi:10.1007/BF02928184

Elliott, A. C., & Hynan, L. S. (2011). A SAS® macro implementation of a multiple comparison post hoc test for a Kruskal–Wallis analysis. *Computer Methods and Programs in Biomedicine*, *102*, 75–80. doi:http://dx.doi.org/10.1016/j.cmpb.2010.11.002

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, *51*, 1–46. doi:10.1111/j.1467-1770.2001.tb00013.x

Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.

Ellis, R. (2005). Principles of instructed language learning. *System*, *33*, 209–224. doi:10.1016/j.system.2004.12.006

Fatemi Jahromi, S. A., & Salimi, F. (2013). Exploring the human element of computer-assisted language learning: an Iranian context. *Computer Assisted Language Learning*, *26*, 158–176. doi:10.1080/09588221.2011.643411

Felix, U. (2003). Teaching languages online : Deconstructing the myths. *Australian Journal of Educational Technology*, *19*(1), 118–138. Retrieved from http://www.ascilite.org.au/ajet/ajet19/felix.html

Felix, U. (2004). A multivariate analysis of secondary students' experience of web-based language acquisition. *ReCALL*, *16*(01), 237–249. doi:10.1017/S0958344004001715

Fios Feasa. (2013). Irish language multimedia and consultancy. Retrieved October 30, 2013, from www.fiosfeasa.com

Fitze, M. (2006). Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning and Technology*, *10*(1), 67–86. Retrieved from www.llt.msu.edu

Foster, P. (2006). Observational research. In R. Sapsford & V. Jupp (Eds.), *Data collection and analysis* (2nd ed., pp. 57–92). London: Sage.

Foster, P., & Snyder Ohta, A. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, *26*, 402–430. doi:10.1093/applin/ami014

Franciosi, S. J. (2011). A comparison of computer game and language-learning task design using flow theory. *CALL-EJ*, *12*(1), 11–25. Retrieved from www.callej.org

Francis, A., & Nusbaum, H. (1999). Evaluating the quality of synthetic speech. In D. Gardner- Bonneau (Ed.), *Human factors and voice interactive systems* (pp. 63–97). Boston: Kluwer Academic Publishers.

Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools. *Language Learning & Technology*, *10*(3), 8–14. Retrieved from http://llt.msu.edu

Gaelscoileanna. (2013). Statistics. Retrieved October 30, 2013, from http://www.gaelscoileanna.ie/en/about/statistics/

Garrett, N. (1991). Technology in the service of language learning: trends and issues. *The Modern Language Journal*, *75*, 74–101. doi:10.1111/j.1540-4781.1991.tb01085.x

Garson, J. (2012). Connectionism. In *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition)*. Retrieved from http://plato.stanford.edu/archives/win2012/entries/connectionism/

Gass, S. M. (2003). Input and interaction. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Oxford: Blackwell Publishing. doi:10.1002/9780470756492.ch9

Gee, J. P. (1992). *The social mind: Language, ideology, and social practice*. New York: Bergin & Garvey.

Gee, J. P. (2005a). Video games, mind and learning. *The Interactive Digital Media & Arts Association Journal*, *2*(1), 37–42.

Gee, J. P. (2005b). Why are video games good for learning? Retrieved October 30, 2013, from http://www.academiccolab.org/resources/documents/MacArthur.pdf

Gee, J. P. (2005c). *Why video games are good for your soul: Pleasure and learning*. Melbourne: Common Ground.

Gee, J. P. (2007). *Good video games + good learning: Collected essays on video games, learning and literacy*. New York: Peter Lang Publisher Inc.

Gee, J. P. (2008a). Game-like learning: An example of situated learning and implications for opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (pp. 200–221). Cambridge: Cambridge University Press.

Gee, J. P. (2008b). Video Games and Embodiment. *Games and Culture*, *3*, 253–263. doi:10.1177/1555412008317309

Gelan, A. (2010). Language and text-to-speech technologies for culture-embedded language learning. In *Proceedings of the International Conference "ICT for Language Learning" 3rd Edition*. Florence, Italy.

Georgila, K., Black, A. W., Sagae, K., & Traum, D. (2012). Practical evaluation of human and synthesized speech for virtual human dialogue systems. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC), 2012*. Istanbul, Turkey.

Gobl, C., Bennett, E., & Ní Chasaide, A. (2002). Expressive synthesis: how crucial is voice quality. In *Proceedings of the IEEE Workshop on Speech Synthesis* (pp. 1–4). Santa Monica, California.

González, D. (2007). Text-to-speech applications used in EFL contexts to enhance pronunciation. *TESL-EJ*, *11*(2), 1–11. Retrieved from http://www.tesl-ej.org/

González-Bueno, M. (1998). The effects of electronic mail on Spanish L2 discourse. *Language Learning and Technology*, *1*(2), 55–70. Retrieved from www.llt.msu.edu

Green, S. B., & Salkind, N. J. (2008). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (5th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.

Greene, C. (2006). Computer-assisted language learning (CALL) for dyslexic students. In K. Miesenberger, J. Klaus, W. L. Zagler, & A. I. Karshmer (Eds.), *Proceedings of the Tenth International Conference Computers Helping People with Special Needs (ICCHP 2006)* (pp. 793–800). Linz, Austria: Springer Berlin Heidelberg. doi:10.1007/11788713_116

Gregg, K. R. (2003). SLA theory: Construction and assessment. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 831–865). Oxford: Blackwell Publishing. doi:10.1002/9780470756492.ch23

Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on*

Educational Resources in Computing, *7*, 1–16. doi:10.1145/1281320.1281322

Gupta, P., & Schulze, M. (2012). Human language technologies (HLT). *Human Language Technologies (HLT). Module 3.5 in Davies, G. (Ed.) Information and Communications Technology for Language Teachers (ICT4LT), Slough, Thames Valley University [Online].* Retrieved October 30, 2013, from http://www.ict4lt.org/en/en_mod3-5.htm

Handley, Z. (2005). *Evaluating text-to-speech (TTS) synthesis for use in computer-assisted language learning (CALL)* (Unpublished doctoral dissertation). University of Manchester, UK.

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, *51*, 906–919. doi:10.1016/j.specom.2008.12.004

Handley, Z., & Hamel, M. J. (2005). Establishing a methodology for benchmarking speech synthesis for computer-assisted language learning (CALL). *Language Learning & Technology*, *9*(3), 99–120. Retrieved from http://llt.msu.edu/

Hanna, N., & Richards, D. (2013). A collaborative agent architecture with human-agent communication model. In F. Dignum, C. Brom, K. Hindriks, M. Beer, & D. Richards (Eds.), *Cognitive Agents for Virtual Environments* (Volume 776., pp. 70–88). London: Springer London, Limited. doi:10.1007/978-3-642-36444-0_5

Harris, J., Forde, P., Archer, P., Nic Fhearaile, S., & O'Gorman, M. (2006). *Irish in primary schools: Long-term national trends in achievement*. Dublin: Stationery Office.

Hashimoto, K., Yamagishi, J., Byrne, W., King, S., & Tokuda, K. (2011). An analysis of machine translation and speech synthesis in speech-to-speech translation system. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*. Prague, Czech Republic: IEEE. doi:10.1109/ICASSP.2011.5947506

Healey, D., & Johnson, N. (1999). *TESOL CALL Interest Section Software List*. Alexandria, VA: Teachers of English to Speakers of Other Languages.

Hecker, L., Burns, L., Elkind, J., Elkind, K., & Katz, L. (2002). Benefits of assistive reading software for students with attention disorders. *Annals of Dyslexia*, *52*, 243–272. doi:10.1007/s11881-002-0015-8

Herrington, J., Oliver, R., & Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Australian Journal of Educational Technology*, *19*(1), 59–71. Retrieved from http://ascilite.org.au/ajet/

Hertz, R. M. (1984). A Software Evaluation Guide For The Language Arts. *CALICO Journal*, *1*(5), 21–23. Retrieved from http://journals.sfu.ca/CALICO/index.php/calico/article/viewFile/209/101

Higgins, J. (1988). *Language, learners, and computers: Human intelligence and artificial unintelligence*. New York: Longman.

Huang, D.-Y. (2011). Prediction of Perceived Sound Quality of Synthetic Speech. In *Proceedings of the Asia-Pacific Signal and Information Processing Association, Annual Summit and Conference 2011*. Xi'an, China.

Huang, X., Acero, A., & Hon, H.-W. (2001). *Spoken language processing: A guide to theory algorithm and system development*. New Jersey: Prentice Hall PTR.

Hubbard, P. (2005). A review of subject characteristics in CALL research. *Computer Assisted Language Learning*, *18*, 351–368. doi:10.1080/09588220500442632

Hubbard, P. (2006). Evaluating CALL software. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching*. San Marcos, TX: Computer Assisted Language Instruction Consortium.

Hubbard, P. (2009). General introduction. In P. Hubbard (Ed.), *Computer Assisted Language Learning, Volume 1: Foundations of CALL, Critical Concepts in Linguistics* (pp. 1–20). New York: Routledge.

Hubbard, P., & Bradin Siskin, C. (2004). Another look at tutorial CALL. *ReCALL*, *16*, 448–461. doi:10.1017/S0958344004001326

Hunt, A. J., & Black, A. W. (1996). Unit selection in a concatenative speech synthesis system using a large speech database. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-96)* (pp. 373–376). Atlanta, GA: IEEE. doi:10.1109/ICASSP.1996.541110

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.

Ipsos MRBI. (2010). *Attitudes towards Irish in school*. Dublin: Comhar na Múinteoirí Gaeilge. Retrieved from http://cnmg.ie/pdfs/10-028750%20%20Comhar%20OmniPoll%20Report%20%25282%2529.pdf

Itahashi, S. (2000). Guidelines for Japanese speech synthesizer evaluation. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC2000)* (pp. 655–660). Athens, Greece.

ITU-T Recommendation P.85. (1994). *A method for subjective performance assessment of the quality of speech output devices*. Geneva. Retrieved from http://www.stat.colostate.edu/~hari/st640/paper1.pdf

Jamieson, J., Chapelle, C., & Preiss, S. (2005). CALL evaluation by developers, a teacher, and students. *CALICO Journal*, *23*(1), 93–138. Retrieved from https://calico.org

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, *38*, 1217–1218. doi:10.1111/j.1365-2929.2004.02012.x

Jancewicz, B., & MacKenzie, M. (2002). Applied computer technology in Cree and Naskapi language programs. *Language Learning & Technology*, *6*(2), 83–91. Retrieved from http://llt.msu.edu/vol6num2/pdf/jancewicz.pdf

Jia, J. (2004). The study of the application of a web-based chatbot system on the teaching of foreign languages. In R. Ferdig et al. (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2004* (pp. 1201–1207). Chesapeake, VA.

Jones, C., & Cross, S. (2009). Is there a net generation coming to university? In *ALT-C 2009 "In Dreams Begins Responsibility": Choice, Evidence and Change, 8-10 September 2009*. Manchester, UK.

Julkunen, K. (2001). Situation- and task- specific motivation in foreign language learning. In Z. Dornyei & R. Schmidt (Eds.), *Motivation and second language learning* (pp. 29–41). Honolulu, HI: University of Hawai'i Press.

Jung, U. O. H. (2005). CALL: past, present and future — a bibliometric approach. *ReCALL*, *17*, 4–17. doi:10.1017/S0958344005000212

Kang, M., Kashiwagi, H., Treviranus, J., & Kaburagi, M. (2008). Synthetic speech in foreign language learning: an evaluation by learners. *International Journal of Speech Technology*, *11*, 97–106. doi:10.1007/s10772-009-9039-3

Kang, M., Kashiwagi, H., Zou, Y., Ohtsuki, K., & Kaburagi, M. (2010). An application of Bayes' theorem to evaluate synthetic speech for computer-assisted language learning. In *Proceedings of the Ninth WSEAS*

International Conference on Education and Educational Technology (pp. 253–257). Iwate, Japan.

Kárpáti, A. (2009). Web 2 technologies for Net Native language learners: a "social CALL." *ReCALL, 21*, 139–156. doi:10.1017/S0958344009000160

Kataoka, H. (2007). Preparation for university entrance examinations: How to learn frequently-appearing English words with the TTS audios. In *Paper presented at the Multimedia & Internet Seminar in 2007, Japan Association for Language Education and Technology (LET)*. Osaka, Japan.

Kataoka, H. (2008). The use of text-to-speech synthesis technology for foreign language education in a Japanese school. In *Proceedings of the Third CLS International Conference* (pp. 318–326). Singapore.

Keller, E., & Zellner-Keller, B. (2000). Speech synthesis in language learning: challenges and opportunities. In *Proceedings of InSTIL 2000 Conference* (pp. 109–116). Dundee, Scotland.

Keller, T., Gerjets, P., Scheiter, K., & Garsoffky, B. (2006). Information visualizations for knowledge acquisition: the impact of dimensionality and color coding. *Computers in Human Behavior, 22*, 43–65. doi:10.1016/j.chb.2005.01.006

Kennedy, G., Dalgarno, B., Gray, K., Judd, T., Waycott, J., Bennett, S., ... Churchward, A. (2007). The net generation are not big users of Web 2.0 technologies: Preliminary findings. In *Proceedings of ASCILITE 2007 Conference* (pp. 517–525). Singapore.

Kerfoot, B. P., Baker, H., Jackson, T. L., Hulbert, W. C., Federman, D. D., Oates, R. D., & DeWolf, W. C. (2006). A multi-institutional randomized controlled trial of adjuvant web-based teaching to medical students. *Academic Medicine, 81*(3), 224–230.

Kickmeier-Rust, M. D., Albert, D., Hockemeyer, C., & Augustin, T. (2007). Not breaking the narrative: Individualized competence assessment in educational games. In *Proceedings of the European Conference on Games-based Learning (ECGBL)* (pp. 161–169). Paisley, Scotland.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model for text comprehension and production. *Psychological Review, 85*(5), 363–394.

Kirriemuir, J., & McFarland, A. (2004). *Literature review in games and learning (Report 8)*. Retrieved from http://telearn.archives-ouvertes.fr/docs/00/19/04/53/PDF/kirriemuir-j-2004-r8.pdf

Kılıçkaya, F. (2006). "Text-to-speech technology": What does It offer to foreign language learners? *CALL-EJ Online*, *7*(2). Retrieved from http://callej.org

Kılıçkaya, F. (2011). Improving pronunciation via accent reduction and text-to-speech software. In M. Levy, F. Blin, C. Bradin Siskin, & O. Takeuchi (Eds.), *WorldCall: International Perspectives on Computer-Assisted Language Learning* (pp. 85–96). New York: Routledge.

Klapper, J. (2003). Taking communication to task? A critical review of recent trends in language teaching. *The Language Learning Journal*, *27*(1), 33–42. doi:10.1080/09571730385200061

Klatt, D. H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, *82*(3), 737–793.

Knill, O., Carlsson, J., Chi, A., & Lezama, M. (2004). *An artificial intelligence experiment in college math education* (p. 9). Cambridge, MA. Retrieved from http://www.math.harvard.edu/~knill/preprints/sofia.pdf

Koul, R. (2003). Synthetic speech perception in individuals with and without disabilities. *Augmentative and Alternative Communication*, *19*, 49–58. doi:10.1080/0743461031000073092

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press Inc.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press Inc.

Krashen, S. D. (1987). *Principles and practice in second language acquisition*. London: Prentice-Hall International.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 263–314). Bingley: Emerald Group.

Lampert, A. (2004). *Evaluation of the MU-Talk speech synthesis system*. ICT Report. Retrieved from sgi.nu/nlp/content/pdf/SynthesisEvaluation.pdf

Leakey, J. (2011). *Evaluating computer-assisted language learning: An integrated approach to effectiveness research in CALL*. Oxford: Peter Lang Publisher Inc.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*, 864–886. doi:10.1016/j.specom.2010.08.014

Lee, E.-J. (2008). Flattery may get computers somewhere, sometimes: The moderating role of output modality, computer gender, and user gender. *International Journal of Human-Computer Studies*, *66*, 789–800. doi:10.1016/j.ijhcs.2008.07.009

Lehtinen, B. (2008). Accessing language using online chatbots. *PeerSpectives Online*, (1), 3–6. Retrieved from http://peerspectives.files.wordpress.com/2012/07/peerspectives-issue-1.pdf

Lemmetty, S. (1999). *Review of speech synthesis technology* (Master's thesis, Helsinki University of Technology). Retrieved from http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/thesis.pdf

Lenhart, A., Kahne, J., Middaugh, E., Rankin Macgill, A., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement. Pew Internet & American Life Project* (pp. 1–64). Washington, DC. Retrieved from http://www.eric.ed.gov/PDFS/ED525058.pdf

Leong, C. K. (1995). Effects of on-line reading and simultaneous DECtalk auding in helping below-average and poor readers comprehend and summarize text. *Learning Disability Quarterly*, *18*(2), 101–116.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.

Levy, M., & Hubbard, P. (2005). Why call CALL "CALL"? *Computer Assisted Language Learning*, *18*, 143–149. doi:10.1080/09588220500208884

Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer assisted language learning*. New Jersey: Lawrence Erlbaum & Associates.

Liddell, P., & Garrett, N. (2004). The new language centers and the role of technology. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 27–40). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55.

Lim, C. P. (2008). Spirit of the game: Empowering students as designers in schools? *British Journal of Educational Technology*, *39*, 996–1003. doi:10.1111/j.1467-8535.2008.00823_1.x

Lim, C. P., & Chai, C. S. (2008). Rethinking classroom-oriented instructional development models to mediate instructional planning in technology-enhanced learning environments. *Teaching and Teacher Education*, *24*, 2002–2013. doi:http://dx.doi.org/10.1016/j.tate.2008.05.004

Little, D. (1994). Words and their properties: Arguments for a lexical approach to pedagogical grammar. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 99–122). Cambridge: Cambridge University Press.

Little, D. (2003). Learner autonomy and second/foreign language learning. *Subject Centre for Languages, Linguistics and Area Studies Good Practice Guide*. Retrieved February 26, 2013, from http://www.llas.ac.uk/resources/gpg/1409

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching*, *1*, 14–29. doi:10.2167/illt040.0

Little, D., Devitt, S., & Singleton, D. (1989). *Learning foreign languages from authentic texts: theory and practice*. Dublin: Authentik, in association with the Centre for Information on Language Teaching and Research (London).

Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge: Cambridge University Press.

Liu, M., Moore, Z., Graham, L., & Lee, S. (2003). A look at the research on computer-based technology use in second language learning : A review of the literature from 1990-2000. *Journal of Research on Technology in Education*, *34*, 250–274. doi:10.1080/15391523.2002.10782348

Livingstone, K. A. (2012). The importance of feedback and reinforcement in computer-assisted language learning (CALL). *Baraton Interdisciplinary Research Journal*, *2*(1), 43–51.

Localization Industry Standards Association. (2008). "Frequently asked questions about LISA and the localization industry." Retrieved October 30, 2013, from http://archive.today/029oZ

Logan, G. D. (1988). Toward and instance theory of automatization. *Psychological Review*, *95*, 492–527. doi:10.1037/0033-295X.95.4.492

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input and second language acquisition* (pp. 268–286). Rowley. MA: Newbury House.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of research on language acquisition: Second language acquisition.* (Volume 2., pp. 413–468). San Diego, CA: Academic Press.

Mac Donnacha, S. (2005). *Staid reatha na scoileanna Gaeltachta 2004: Léargas ó na príomhoidí, na múinteoirí agus na tuismitheoirí* (pp. 1–26). Baile Átha Cliath: An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta.

Mac Donnacha, S., Ní Chualáin, F., Ní Shéaghdha, A., & Ní Mhainín, T. (2005). *Staid reatha na scoileanna Gaeltachta* (pp. 1–19). Baile Átha Cliath: An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta.

Mac Giolla Chríost, D. (2012). A question of national identity or minority rights? The changing status of the Irish language in Ireland since 1922. *Nations and Nationalism*, *18*, 398–416. doi:10.1111/j.1469-8129.2011.00508.x

Mac Gréil, M., & Rhatigan, F. (2009). *The Irish language and the Irish people: Report on the attitudes towards, competence in and use of the Irish language in the Republic of Ireland in 2007-08* (pp. 1–126). National University of Ireland Maynooth. Retrieved from http://www.mayococo.ie/en/Services/OifignaGaeilge/Publications

Mannell, R. H. (1984). Natural and synthetic speech intelligibility and quality testing. In *Aspects of Speech Synthesis Performance (Unpublished Honours dissertation)* (Macquarie., pp. 39–61). Retrieved from http://clas.mq.edu.au/synthesis/synth_evaluation/intelqual.html

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52. doi:10.1207/S15326985EP3801_6

Mayo, C., Clark, R. A. J., & King, S. (2005). Multidimensional scaling of listener responses to synthetic speech. In *Proceedings of Interspeech 2005* (pp. 1725–1728). Lisbon, Portugal.

Mayo, C., Clark, R. A. J., & King, S. (2011). Listeners' weighting of acoustic cues to synthetic speech naturalness: A multidimensional scaling analysis. *Speech Communication*, *53*, 311–326. doi:10.1016/j.specom.2010.10.002

McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). *A Literature Review of Gaming in Education* (pp. 1–23). Retrieved from http://www.pearsonassessments.com/hai/Images/tmrs/Lit_Review_of_Gaming_in_Education.pdf

McCosker, C. (2009). Operation conjugation! A game-based learning approach to language studies. In *Paper presented at EdTech09; Tenth Annual Irish Educational Technology Users Conference*. Dublin: National College of Ireland. Retrieved from http://ilta.ie/edtech/edtech-2009-2/edtech-2009-postgraduate-presentations/

McLeod, S. A. (2008). Likert scale. *SimplyPsychology*. Retrieved October 30, 2013, from http://www.simplypsychology.org/likert-scale.html

Miceli, T., Murray, S. V., & Kennedy, C. (2010). Using an L2 blog to enhance learners' participation and sense of community. *Computer Assisted Language Learning*, *23*(4), 321–341. doi:10.1080/09588221.2010.495321

Mikk, J. (2008). Sentence length for revealing the cognitive load reversal effect in text comprehension. *Educational Studies*, *34*, 119–127. doi:10.1080/03055690701811164

Mishan, F. (2005). *Designing authenticity into language learning materials*. Bristol: Intellect Books.

Mitchell, A., & Savill-Smith, C. (2004). *The use of computer and video games for learning: A review of the literature*. London: Learning and Skills Development Agency.

Mitchell, B. L. (2012). *Game design essentials*. Hoboken, New Jersey: John Wiley & Sons.

Mitgutsch, K., & Alvarado, N. (2012). Purposeful by design? A serious game design assessment framework. In *Proceedings of the International Conference on the Foundations of Digital Games (FDG '12)* (pp. 121–128). New York, USA: ACM. doi:10.1145/2282338.2282364

Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities*, *29*(3), 271–279. doi:10.1177/002221949602900305

Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior*, *24*, 2530–2540. doi:10.1016/j.chb.2008.03.012

Morton, H., Gunson, N., Marshall, D., McInnes, F., Ayres, A., & Jack, M. (2011). Usability assessment of text-to-speech synthesis for additional detail in an automated telephone banking system. *Computer Speech & Language*, *25*, 341–362. doi:10.1016/j.csl.2010.05.008

Moseley, C. (Ed.). (2010). *Atlas of the world's languages in danger* (3rd ed.). Paris: UNESCO Publishing. Retrieved from http://www.unesco.org/culture/en/endangeredlanguages/atlas

Murray, L., & Hourigan, T. (2008). Blogs for specific purposes: Expressivist or socio-cognitivist approach? *ReCALL*, *20*, 83–98. doi:10.1017/S0958344008000719

Murtagh, L. (2003). *Retention and attrition of Irish as a second language: A longitudinal study of general and communicative proficiency in Irish among second level school leavers and the influence of instructional background, language use and attitude/motivation variable* (Doctoral dissertation, University of Groningen, Amsterdam). Retrieved from http://www.let.rug.nl/languageattrition/dissertations

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*, 171–181. doi:10.1037//1076-898X.7.3.171

Navas, E., Hernaez, I., & Sanchez, J. M. (2002). Subjective evaluation of synthetic intonation. In *Proceedings of the 2002 IEEE Workshop on Speech Synthesis* (pp. 23–26). Santa Monica, CA. doi:10.1109/WSS.2002.1224364

Ní Chasaide, A. (1999). Illustrations of the IPA. Irish. In *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet* (pp. 111–116). Cambridge: Cambridge, Cambridge University Press.

Ní Chasaide, A., Ní Chiaráin, N., Wendler, C., Berthelsen, H., Kelly, A., Gilmartin, E., … Gobl, C. (2011). Towards personalised, synthesis-based content in Irish (Gaelic) language education. In *Proceedings of the ISCA Special Interest Group link (SIG) on Speech and Language Technology in Education (SLaTE)* (pp. 29–32). Venice, Italy.

Ní Chasaide, A., Wogan, J., Ó Raghallaigh, B., Ní Bhriain, Á., Zoerner, E., Berthelsen, H., & Gobl, C. (2006). Speech technology for minority languages: the case of Irish (Gaelic). In *Proceedings of the 9th International Conference on Spoken Language Processing* (pp. 181–184). Pittsburgh, PA.

Ní Chiaráin, N., & Ní Chasaide, A. (2012). Exploring Irish text-to-speech synthesis for educational purposes. In *Proceedings of the Workshop on Innovation and Applications in Speech Technology (IAST)* (p. 37). Dublin, Ireland.

Noble, D. F. (2001). The learning curve. Online special. *The Digital Degree, ABC Radio National Broadcast. Part two: Dancing with the Devil.* Retrieved August 30, 2012, from http://www.abc.net.au/radionational/programs/bigideas/the-digital-degree---part-two-dancing-with-the/3491734

Ó Baoill, D. (1981). *Earráidí scríofa Gaeilge. Cuid 3, réamhfhocail agus comhréir : earráidí a tharla in aistí Gaeilge na hÁrdteistiméireachta, 1975.* Baile Átha Cliath: Institiúd Teangeolaíochta Éireann.

Ó Baoill, D. (1987). *Lárchanúint don Ghaeilge*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.

Ó Domhnalláin, T., & Ó Baoill, D. (1978). *Earráidí scríofa Gaeilge. Cuid 1, earráidí briathra: cuntas ar na hearráidí Gaeilge a rinne sampla ionadaíoch de na hiarrthóiri ar scrúdú Gaeilge na hArdteistiméireachta*. Baile Átha Cliath: Institiúd Teangeolaíochta Éireann.

Ó Domhnalláin, T., & Ó Baoill, D. (1979). *Earráidí scríofa Gaeilge. Cuid 2, ainmfhocail, cáilitheoirí, forainmneacha, cónaisc agus míreanna*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.

Ó Fathaigh, M., & Ní Bhraoin, A. (1996). *Irish language attitudes, competence & usage among University College Cork staff: Some empirical findings*. Corcaigh: Bord na Gaeilge, Coláiste na hOllscoile, Corcaigh.

Ó Flatharta, P. (1989). The teaching of Irish as a first language in primary schools in the Gaeltacht. *Teanga*, *9*, 75–83.

Ó Ghadhra, N. (2001). The Gaeltacht and the future of Irish. *Studies: An Irish Quarterly Review*, *90*(360), 433–444.

Ó Giollagáin, C., Mac Donnacha, S., Ní Chualáin, F., Ní Shéaghdha, A., & O'Brien, M. (2007). *Comprehensive linguistic study of the use of Irish in the Gaeltacht: Principal findings and recommendations*. Acadamh na hOllscolaíochta Gaeilge, National University of Ireland, Galway; National University of Ireland, Maynooth. doi:ISBN: 0-7557-7334-9

Ó Murchú, H. (2001). *Irish: The Irish language in education in the Republic of Ireland*. Ljouwert/Leeuwarden, The Netherlands: Mercator European Research Centre on Multilingualism and Language Learning.

Ó Raghallaigh, B. (2009). *Multi-dialect phonetisation for Irish text-to-speech synthesis: a modular approach* (Unpublished doctoral dissertation). Trinity College, Dublin.

Ó Raghallaigh, B. (2012). *Fuaimeanna na Gaeilge*. Baile Átha Cliath: Cois Life Teoranta.

Ó Riagáin, P., & Ó Gliasáin, M. (1994). *The national surveys on languages 1993: Preliminary report*. Baile Átha Cliath: Institiúd Teangeolaíochta Éireann.

Ó Riain, S. (2010). Irish and translation - the EU context. *Études Irlandaises*, *35*(2), 65–80. Retrieved from http://etudesirlandaises.revues.org/1958

Ó Siadhail, M. (1980). *Learning Irish*. Dublin: Dublin Institute for Advanced Studies.

O'Beirne Ranelagh, J. (2012). *A short history of Ireland* (3rd ed.). New York: Cambridge University Press.

O'Donnell, E., & Sharp, M. (2012). Students' views of e-learning: The impact of technologies on learning in higher education in Ireland. In K. Moyle & G. Wijngaards (Eds.), *Student Reactions to Learning with Technologies: Perceptions and Outcomes 2012* (pp. 204–225). University of Canberra, Australia, and Holland University, The Netherlands. doi:10.4018/978-1-61350-177-1.ch010

O'Dowd, J. (2012). Irish-medium schools. *Northern Ireland Assembly Debates*. Retrieved from http://www.theyworkforyou.com/ni/?id=2012-02-20.0.136

O'Sullivan, C., & Ennis, C. (2011). Metropolis: multisensory simulation of a populated city. In *Proceedings of the Third International Conference on Games and Virtual Worlds for Serious Applications* (pp. 1–7). Athens, Greece: IEEE Computer Society. doi:10.1109/VS-GAMES.2011.9

Olson, R. K., & Wise, B. W. (1992). Reading on the computer with orthographic and speech feedback. *Reading and Writing*, *4*, 107–144. doi:10.1007/BF01027488

Osuna, M. M., & Meskill, C. (1998). Using the world wide web to integrate Spanish language and culture: A pilot study. *Language Learning and Technology*, *1*(2), 71–90. Retrieved from www.llt.msu.edu

Otto, S. E. K., & Hatasa, K. (2006). Teaching technologies: Second language. In *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/00616-7

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. doi:10.1037/h0084295

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead: Open University Press.

Pandorabots. (2013). Pandorabots: A multilingual chatbot hosting service. Retrieved October 30, 2013, from http://www.pandorabots.com/

Parker, L. L. (2005). *Language development technologies for young English learners*. University of California, Office of the President, California. Retrieved from http://www.ucop.edu/elltech/parkerpaper1-21-05.pdf

Peirce, N., & Wade, V. (2010). Personalised learning for casual games: The "Language Trap" online language learning. In *Proceedings of the Fourth European Conference on Game Based Learning (ECGBL 2010)* (pp. 306–315). Copenhagen, Denmark: Bente Meyer.

Pellegrini, T., Costa, Â., & Trancoso, I. (2012). Less errors with TTS? A dictation experiment with foreign language learners. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech 2012)*. Portland, Oregon: ISCA.

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming*, *41*, 71–93. doi:10.1177/1046878109355684

Pica, T. (1997). Second language teaching and research relationships: a North American view. *Language Teaching Researcch*, *1*(1), 48–72. doi:10.1177/136216889700100104

Pillay, H. K. (1994). Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science*, *22*, 91–113. doi:10.1007/BF00892159

Podsiadlo, M. (2007). *Large scale speech synthesis evaluation (*Master's thesis, University of Edinburgh, Scotland). Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/1943

Popham, W. J. (2008). All about assessment / A misunderstood grail. *Educational Leadership*, *66*(1), 82–83.

Pourhossein Gilakjani, A. (2012). Visual, auditory, kinaesthetic learning styles and their impacts on English language teaching. *Journal of Studies in Education*, *2*, 104–113. doi:10.5296/jse.v2i1.1007

Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 20). Oxford: Oxford University Press.

Prensky, M. (2007). *Digital game-based learning*. St. Paul, Minnesota: Paragon House.

Pressey, S. L. (1926). A simple apparatus which gives tests – and teaches. *School and Society*, *23*(586), 373–376.

Prys, D., Williams, B., Hicks, B., Jones, D., Ní Chasaide, A., Gobl, C., ... Uí Dhonnchadha, E. (2004). WISPR: Speech processing resources for Welsh and Irish. In *Proceedings of the SALTMIL Workshop at LREC2004: First Steps in Language documentation for Minority Languages* (pp. 68–71). Lisbon, Portugal.

Purushotma, R., Thorne, S. L., & Wheatley, J. (2009). 10 key principles for designing video games for foreign language learning. *Open Language & Learning Games Project, Massachusetts Institute of Technology Education Arcade*. Retrieved October 30, 2013, from http://lingualgames.wordpress.com/article/10-key-principles-for-designing-video-27mkxqba7b13d-2/

Raizen, E., & Lippmann, J. (2004). TTS tools and language instruction. *Foreign Language Teachers' Toolbox*.

Ratan, R., & Ritterfeld, U. (2009). Classifying serious games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 10–22). New York: Routledge.

Reeder, K., Heift, T., Roche, J., Tabyanian, S., Schlickau, S., & Gölz, P. (2004). Toward a theory of EValuation for second language learning media. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 255–278). Mahwah, NJ: Lawrence Erlbaum Associates.

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge: Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge: Cambridge University Press.

Ritterfeld, U., Cody, M., & Vorderer, P. (2009). Serious games-explication of an oxymoron-introduction. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects*. (pp. 3–9). New York: Routledge.

Robin, R. (2007). Commentary: Learner-based listening and technological authenticity. *Language Learning & Technology*, *11*(1), 109–115. Retrieved from http://llt.msu.edu

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 631–677). Oxford: Blackwell.

Rudo, P. (2011). Real-world artificial intelligence has arrived: A chatbot can automate your customer service, support and interactions. *Enterprise Features*. Retrieved February 27, 2013, from http://enterprisefeatures.com/2011/05/real-world-artificial-intelligence-has-arrived-a-chatbot-can-automate-your-customer-service-support-and-interactions/

Ruxton, G. D., & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, *19*, 690–693. doi:10.1093/beheco/arn020

Salaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal*, *85*, 39–56. doi:10.1111/0026-7902.00096

Sampson, G. (1980). *Schools of linguistics*. Stanford, California: Stanford University Press.

Sasaki, A. (2006). Text-to-speech on digital flashcards and its effect on vocabulary learning. In *Proceedings of the 46th Annual Conference of Language Education and Technology (LET)* (pp. 451–454).

Savignon, S. J. (2006). Communicative language teaching. In *Encyclopedia of Language & Linguistics*. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/00610-6

Scholfield, P. J. (2003). *Evaluation of CALL software*. Retrieved from http://privatewww.essex.ac.uk/~scholp/calleval.htm

Schrooten, W. (2006). Task-based language teaching and ICT: Developing and assessing interactive multimedia for task-based language teaching. In K. Van den Branden (Ed.), *Task-Based Language Education: From theory to practice* (pp. 129 – 150). Cambridge: Cambridge University Press.

SecondLife. (2011). Linden dollar economy metrics up, users and usage unchanged. *Community SecondLife*. Retrieved October 30, 2013, from http://community.secondlife.com/t5/Featured-News/Q1-2011-Linden-Dollar-Economy-Metrics-Up-Users-and-Usage/ba-p/856693

Sha, G. (2010). Using TTS voices to develop audio materials for listening comprehension: A digital approach. *British Journal of Educational Technology, 41*, 632–641. doi:10.1111/j.1467-8535.2009.01025.x

Shi, C. (2011). A study of the relationship between cognitive styles and learning strategies. *Higher Education Studies, 1*(1), 20. doi:10.5539/hes.v1n1p20

Silberman, C. E. (1970). *Crisis in the classroom: the remaking of American education*. New York: Random House.

Sityaev, D., Knill, K., & Burrows, T. (2006). Comparison of the ITU-T P.85 standard to other methods for the evaluation of text-to-speech systems. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006)* (pp. 1077 – 1080). Pittsburgh, PA.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory* (pp. 133–171). New York: John Wiley & Sons.

Squires, D., & McDougall, A. (1996). Software evaluation: a situated approach. *Journal of Computer Assisted Learning, 12*, 146–161. doi:10.1111/j.1365-2729.1996.tb00047.x

State Examinations Commission. (2011). *Annual report 2011*. Retrieved from http://www.examinations.ie/about/SEC_Annual_Report_2011_Final_English_Version.pdf

Stern, S. E., Mullennix, J. W., & Yaroslavsky, I. (2006). Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies, 64*, 43–52. doi:10.1016/j.ijhcs.2005.07.002

Stevens, C., Lees, N., Vonwiller, J., & Burnham, D. (2005). On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference. *Computer Speech & Language, 19*, 129–146. doi:http://dx.doi.org/10.1016/j.csl.2004.03.003

Stevension, A., & Soanes, C. (Eds.). (2003). The Oxford Dictionary of English. Oxford: Clarendon Press.

Stratil, M., Burkhardt, D., Jarratt, P., & Yandle, J. (1987). Computer-aided language learning with speech synthesis: User reactions. *Innovations in Education & Training International*, *24*, 309–316. doi:10.1080/0033039870240408

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious games – An overview* (pp. 1–24). Skövde, Sweden. Retrieved from http://www.autzones.com/din6000/textes/semaine12/SusiEtAl(2005).pdf

Sweller, J. (2005). The Redundancy Principle. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 159–168). New York: Cambridge University Press.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. Retrieved from http://llt.msu.edu

Sykes, J. M. (2008). *A dynamic approach to social interaction: SCMC, synthetic immersive environments & Spanish pragmatics* (Doctoral dissertation, University of Minnesota). Retrieved from http://conservancy.umn.edu/bitstream/91603/1/Sykes_Julie May 2008.pdf

Sykes, J. M., Oskoz, A., & Thorne, S. L. (2008). Web 2.0, synthetic immersive environments, and mobile resources for language education. *CALICO Journal*, *25*, 528–546. doi:10.11139/cj.25.3.528-546

Tam, S. S., & Taher, B. (2012). Learner-centred approach to project-based computer-assisted language learning. *MJAL*, *4*(3), 129–140.

Team Engage. (2009). 8D taps language learners, bots, microtransactions. *engagedigital.com*. Retrieved October 20, 2013, from http://www.engagedigital.com/blog/2009/05/29/out-of-stealth-8d-taps-language-learners-bots-microtransactions/

Thomas, L. (1997). Retrospective power analysis. *Conservation Biology*, *11*, 276–280. doi:10.1046/j.1523-1739.1997.96102.x

Toda, T., & Tokuda, K. (2007). A speech parameter generation algorithm considering global variance for HMM-based speech synthesis. *EICE TRANSACTIONS on Information and Systems*, *90*(5), 816–824.

Tong, J., & Cadwell-Harris, C. (2008). *Emotions & language acquisition: How to help an adult learn a second language* (pp. 1–4). Boston. Retrieved from bupsychling.files.wordpress.com/2009/07/videogame.doc

Uí Dhonnchadha, E. (2010). *Natural language processing tools: Developing a part-of-speech tagger and partial dependency parser for Irish*. Saarbrücken: LAP Lambert Academic Publishing.

Underwood, J. (1984). *Linguistics, computers and the language teacher: a communicative approach*. Rowley, MA: Newbury House.

Valentini-Botinhao, C., Yamagishi, J., & King, S. (2011). Evaluation of objective measures for intelligibility prediction of HMM-based synthetic speech in noise. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)* (pp. 5112–5115). Prague, Czech Republic.

Van Bezooijen, R., & van Heuven, V. J. (1997). Assesment of synthesis systems. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of standards and resources for spoken language systems* (Volume 3., pp. 481– 563). New York: Mouton de Gruyter.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, *41*(2), 16–30. Retrieved from http://www.educause.edu/

Van Lun, E., Verduijn, X., Kuligowska, K., de Wolf, A., & Morton, D. (2013). chatbots.org. Retrieved October 31, 2013, from www.chatbots.org

Van Merrienboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147–177. doi:10.1007/s10648-005-3951-0

Van Schaik, P., & Ling, J. (2003). The effect of link colour on information retrieval in educational intranet use. *Computers in Human Behavior*, *19*, 553–564. doi:10.1016/S0747-5632(03)00004-9

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of Cascade. *The Journal of the Learning Sciences*, *8*, 71–125. doi:10.1207/s15327809jls0801_3

Veaux, C., Yamagishi, J., & King, S. (2012). Using HMM-based speech synthesis to reconstruct the voice of individuals with degenerative speech disorders. In *Proceedings of Interspeech 2012*. Portland, Oregon.

Viswanathan, M., & Viswanathan, M. (2005). Measuring speech quality for text-to-speech systems: Development and assessment of a modified mean opinion score (MOS) scale. *Computer Speech & Language*, *19*, 55–83. doi:10.1016/j.csl.2003.12.001

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), *Speech Intelligibility and Speaker Recognition*. Stroudsburg. PA: Dowden, Hutchinson and Ross.

Walsh, J. (2012). Language policy and language governance: a case-study of Irish language legislation. *Language Policy*, *11*, 323–341. doi:10.1007/s10993-012-9238-7

Wang, Y. (2008). *Designing chatbot interfaces for language learning: Ethnographic research into affect and users' experiences* (Unpublished master's thesis). The University of British Columbia, Vancouver.

Ward, M. (2001). *A template for CALL programs for endangered languages* (Master's thesis, Dublin City University). Retrieved from http://www.computing.dcu.ie/~mward/mthesis.html

Ward, M. (2004). The additional uses of CALL in the endangered language context. *ReCALL*, *16*, 345–359. doi:10.1017/S0958344004000722

Ward, M. (2007). *The integration of CL resources in CALL for Irish in the primary school context*. Dublin City University. Retrieved from http://doras.dcu.ie/16995/1/monica_ward_SC.pdf

Warren, S. J., Dondlinger, M. J., & Barab, S. A. (2008). A MUVE towards PBL writing: Effects of a digital learning environment designed to improve elementary school writing. *Journal of Research on Technology in Education*, *41*(1), 113–140.

Warschauer, M., & Healey, D. (1998). Computers and language learning: an overview. *Language Teaching*, *31*, 57–71. doi:10.1017/S0261444800012970

Weizenbaum, J. (1966). ELIZA- A computer program for the study of natural language communication between men and machine. *Communications of the ACM*, *9*(1), 36–45.

White, L. (2003). On the nature of interlanguage representation: Universal grammar in the L2. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 19–42). Malden, MA: Blackwell Publishing. doi:10.1002/9780470756492.ch2

Wickens, C. D. (2007). Attention to attention and its applications: A concluding view. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From Theory to Practice*. New York: Oxford University Press.

Williams, B., Prys, D., & Ní Chasaide, A. (2005). Creating an ongoing research capability in speech technology for two minority languages: experiences

from the WISPR project. In *Proceedings of the Ninth European Conference on Speech Communication and Technology, (INTERSPEECH 2005)* (pp. 189–192). Lisbon, Portugal.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86. Retrieved from http://llt.msu.edu

Wogan, J., Ó Raghallaigh, B., Ní Bhriain, Á., Zoerner, E., Berthelsen, H., Ní Chasaide, A., & Gobl, C. (2006). Developing a spoken corpus and a synthesiser for Irish (Gaelic). In *Proceedings of the Fifth SALTMIL Workshop on Minority Languages*. Genoa, Italy.

Worcester, R. M., & Burns, T. R. (1975). A statistical examination of the relative precision of verbal scales. *Journal of Market Research Society, 17*(3), 181–197.

Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K., & Isogai, J. (2009). Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm. *IEEE Transactions on Audio, Speech and Language Processing, 17*(1), 66–83.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology, 13*(2), 48–67. Retrieved from http://llt.msu.edu

Yoshida, S. (2008). Development and practice of an electronic phrasal verb wordbook with GIF animations. In *Proceedings of the Third International Conference of WorldCALL* (pp. 93–97). Fukuoka, Japan. Retrieved from http://www.j-let.org/~wcf/proceedings/proceedings.pdf

Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., & Kitamura, T. (1999). Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis. In *Proceedings of Eurospeech 1999, Vol. 4* (pp. 1623–1626). Budapest, Hungary.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., ... Woodland, P. (2009). *The HTK book (for HTK version 3.4)*. Cambridge University Engineering Department,. Retrieved from http://htk.eng.cam.ac.uk/

Yugandhar, K., Srinivas, V., & Ramakrishna Rao, A. (2010). Conveniences and complexities of computer-aided language learning. *Language in India, 10*(1), 360–367. Retrieved from http://www.languageinindia.com

Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A. W., & Tokuda, K. (2007). The HMM-based speech synthesis system (HTS) version 2.0. In *Proceedings of the Sixth ISCA Workshop on Speech Synthesis* (pp. 294–299). Bonn, Germany.

Zhang, M., Tao, J., Jia, H., & Wang, X. (2008). Improving HMM based speech synthesis by reducing over-smoothing problems. In *Proceedings of the Sixth International Symposium on Chinese Spoken Language Processing (ISCSLP '08)* (pp. 1–4). Kunming, China.

Zheng, D., Young, M. F., Wagner, M. M., & Brewer, R. A. (2009). Negotiation for action: English language learning in game-based virtual worlds. *The Modern Language Journal*, *93*, 489–511. doi:10.1111/j.1540-4781.2009.00927.x

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology*, *40*, 790–803. doi:10.1111/j.1467-8535.2008.00859.x