Phonotactic Probability and Phonotactic Constraints: Processing and Lexical Segmentation by Arabic Learners of English as a Foreign Language

A thesis submitted to the Faculty of Humanities and Social Sciences of
Newcastle University in fulfillment of the requirements for the PhD degree in
Applied Linguistics

Faisal M. A. Al-jasser

2008

Abstract

A fundamental skill in listening comprehension is the ability to recognize words. The ability to accurately locate word boundaries (i.e. to lexically segment) is an important contributor to this skill. Research has shown that English native speakers use various cues in the signal in lexical segmentation. One such cue is phonotactic constraints; more specifically, the presence of illegal English consonant sequences such as /tl/ and /dl/ signals word boundaries. It has also been shown that phonotactic probability (i.e. the frequency of segments and sequences of segments in words) affects native speakers' processing of English. However, the role that phonotactic probability and phonotactic constraints play in the EFL classroom has hardly been studied, while much attention has been devoted to teaching listening comprehension in EFL.

This thesis reports on an intervention study which investigated the effect of teaching English phonotactics upon Arabic speakers' lexical segmentation of running speech in English. The study involved a native English group (N= 12), a non-native speaking control group (N= 20); and a non-native speaking experimental group (N=20). Each of the groups took three tests, namely Non-word Rating, Lexical Decision and Word Spotting. These tests probed how sensitive the subjects were to English phonotactic probability and to the presence of illegal sequences of phonemes in English and investigated whether they used these sequences in the lexical segmentation of English. The non-native groups were post-tested with the same tasks after only the experimental group had been given a treatment which consisted of explicit teaching of relevant English phonotactic constraints and related activities for 8 weeks. The gains made by the experimental group are discussed, with implications for teaching both pronunciation and listening comprehension in an EFL setting.

Acknowledgements

It would not have been possible to complete this thesis without the appreciated support of many great people. Heartfelt thanks first go to my main supervisor, Dr Martha Young-Scholten and to my second supervisor, Dr Ghada Kattab. From Martha I learned what good supervision is. It is not to dictate what should be made, but rather to make sure that the steps the student takes are the correct ones. Her guidance, support, understanding, advice and, above all, belief in me were invaluable. I am also grateful to her for kindly volunteering to record the test materials used in the experiments reported in the present thesis. Given my initial research proposal, Martha may have never expected that I would end up doing what I did, but I respect her deeply for sticking with it as we found a way to unite our interests. I have been really privileged to have such a dedicated scholar as my supervisor.

I have also been equally privileged to have Dr Ghada Khattab as my second supervisor. Although she did not officially have to, Ghada has been involved in all aspects of my research. She taught me most of the technical skills related to the creation of my stimuli items, provided access to some native subjects participating in the tasks and responded to my numerous questioning emails immediately. Importantly, she critically commented on my work at all phases. Ghada, I sincerely appreciate it.

I am also thankful for other people who contributed directly or indirectly to the completion of my research. I am deeply indebted to Dr Michael Vitevitch from the University of Kansas, USA for kindly calculating the neighbourhood density of my non-word stimuli items. My thanks also go to Dr Chris Letts for installing the DMDX on my own laptop, creating the required script and providing technical advice. Dr Simon Kometa and Adil Widyan kindly provided statistical advice. I am also grateful to Dr Thomas Klee from Newcastle University and Dr Melinda Whong-Barr from Leeds University for taking part in my study. My sisters Wafa, Iman, Intisar and my wife Minirah kindly helped me input my data while I was doing my fieldwork in Saudi Arabia. In many aspects of presentation and rewording, I benefited enormously from feedback from Dr John Field, the guest editor of System Journal, and two anonymous readers on a paper of mine, drawing on this thesis, which appeared in a special issue of System journal in March 2008. I also benefited from stimulating discussions with Professor Ellen Broselow, Professor Noel Burton Roberts, Dr S.J. Hannahs, Professor Anders Holmberg and Dr Geoffrey Poole.

For providing the financial support that made this work possible, I would like to thank my sponsor, Qassim University, Saudi Arabia. I am deeply appreciative of the help that colleagues and students in the Department of English Language and Translation, Qassim University provided to make it possible to conduct my experimental study. Due thanks also go to all the staff in the Saudi Cultural Bureau, London, particularly my advisors, Mr Lateef Ibraheem and Mr Mohamed El-ansari for their advice.

Table of Contents

Abstract i
Acknowledgements ii
Table of Contents iv
List of Figures vii
List of Tables x
Chapter 1: Introduction 1
1.1 The problem 1
1.1.1 Connected speech 2
1.1.2 Lexical segmentation 3
1.2 Phonotactics of English and Arabic 6
1.3 Context and significance of the study 7
1.4 Overview of the thesis 10
Chapter 2: L2 Listening Comprehension: Auditory Processing and Teaching 13
2.1 Introduction 13
2.2 Listening comprehension 15
2.2.1 Memory and listening comprehension 17
2.2.2 Controlled vs. automatic processing. 23
2.2.3 Top-down and bottom-up processing 30
2.3 Teaching/training L2 listening 33
2.3.1 Introduction 33
2.3.2 Teaching top-down 36
2.3.3 Teaching bottom-up 51
2.4 Summary 61
Chapter 3: Lexical Segmentation 63
3.1 Introduction 63
3.2 Serendipitous Segmentation Models 66
3.2.1 Summary 74
3.3 Explicit Segmentation Models 76
3.3.1 Allophonic variations 77
3.3.2 Prosodic Cues 83
3.3.3 Phonotactic restrictions 88
3.3.4 How are phonotactics learnt? 103
3.3.5 Language-specific cues 108
3.4 Teaching Phonotactics 116
3.5 Phonotactic Probability 118
3.6 Summary 125
Chapter 4: Methodology 128
4.1 Introduction 128
4.2 Subjects 131
4.2.1 Native English speaking control group 132
4.2.2 Non-native groups 133
4.3 Design 135
4.4 Task overview 136
4.4.1 Non-Word Rating 137
4.4.1.1 Materials for Non-Word Rating Task 138
4.4.1.2 Procedure 144
4.4.1.3 Research questions and hypotheses 147
4.4.2 Auditory Lexical Decision 148
4.4.2.1 Materials for the Lexical Decision Task 150

4.4.2.2 Procedure 153
4.4.2.3 Research questions and hypotheses 154
4.4.3 Word Spotting Task 156
4.4.3.1 Materials for the Word Spotting Task 157
4.4.3.2 Procedure 161
4.4.3.3 Research questions and hypotheses 162
4.4.4 Piloting 163
4.4.5 Treatment 165
4.4.6 Post-tests 169
4.4.6.1 Hypotheses 169
4.7 Summary 170
Chapter 5: Results and Discussion 172
5.1 Introduction 172
5.2 Pre-test results 172
5.2.1 Results of the Non-word Rating Task 172
5.2.1.1 Native Control Group 174
5.2.1.2 Comparing non-native groups (between groups comparison) 175
5.2.1.3 Comparing Native Control to Non-Native Groups 177
5.2.1.4 Non-native groups (within groups comparison) 179
5.2.1.5 Discussion 180
5.2.2 Results of the Lexical Decision Task 191
5.2.2.1 Native control 193
5.2.2.2 Comparing non-native groups (between groups comparison) 194
5.2.2.3 Comparing Native Control to Non-Native Groups 196
5.2.2.4 Non-native groups (within groups comparison) 197
5.2.2.5 Discussion 198
5.2.3 Results of the Word Spotting Task 208
5.2.3.1 Native control 210
5.2.3.2 Comparing non-native groups (between groups comparison) 211
5.2.3.3 Comparing Native Control to Non-Native Groups 213
5.2.3.4 Non-native groups (within groups comparison) 215
5.2.3.5 Discussion 216
5.3 Post-tests' results: the effect of the treatment 220
5.3.1 Results of the Non-word Rating Task 220
5.3.1.1 Comparing non-native control pre-test with post-test results 221
5.3.1.2 Comparing native control to non-native control post-test results 222
5.3.1.3 Comparing non-native experimental pre-test with post-test results 223
5.3.1.4 Comparing native control to non-native experimental post-test results 225
5.3.1.5 Discussion 226
5.3.2 Results of the Lexical Decision Task 230
5.3.2.1 Comparing non-native control pre-test with post-test's results 230
5.3.2.2 Comparing native control to non-native control post-test's results 232
5.3.2.3 Comparing non-native experimental pre-test with post-test's results 233
5.3.2.4 Comparing native control to non-native experimental post-test results 235
5.3.2.5 Discussion 237
5.3.3 Results of the Word Spotting Task 241
5.3.3.1 Comparing non-native control pre-test with post-test's results 241
5.3.3.2 Comparing native control to non-native control post-test's results 243
5.3.3.3 Comparing non-native experimental pre-test with post-test's results 245

5.3.3.4 Comparing native control to non-native experimental post-test's results 246
5.3.2.5 Discussion 248
5.4 Summary 250
Chapter 6: Conclusion and Pedagogical implications 254
6.1 Major findings 254
6.2 Limitations 255
6.3 Pedagogical implications 256
6.3.1 Extra-classroom exposure 256
6.3.2 Awareness raising activities 257
6.3.3 Vocabulary teaching 259
6.4 Suggestions for further research 260
6.5 Conclusion 263
References 264
Appendix A.1 Consent and a survey filled in by non-native subjects prior to participating in the study 283
Appendix A.2 The thirteen non-words starting with consonant clusters that were used to judge which clusters are legal in Qassimi Arabic. 284
Appendix A.3 Transcription of the 6 non-words starting with illegal consonant clusters in both English and Qassimi Arabic with the sum of phoneme and biphone probabilities for the syllable starting at the second consonant of the non-word: 284
Appendix A.4 Transcription of the 6 non-words starting with illegal consonant clusters only in English with the sum of phoneme and biphone probabilities for the syllable starting at the second consonant of the non-word: 284
Appendix A.5 Statistics of non-words with high phonotactic probability: 285
Appendix A.6 Statistics of non-words with low phonotactic probability 285
Appendix A.7 20 Real-word fillers used in the Word Rating and Lexical Decision Tasks. 286
Appendix A.8 Questionnaire used in the Word Rating Task 286
Appendix A.9 A complete list of the items in the order they appeared in the Word Rating Task. 289
Appendix A.10 Four additional real-word fillers used in the Lexical Decision Task. 290
Appendix A.11 three trial items used in the Lexical Decision Task. 290
Appendix B. Items used in the word Spotting task 290
Appendix B.1 Practice items used before each list. 290
Appendix B.2 10 real-word bearing fillers. 291
Appendix B.3 14 Bisyllabic nonsense sequences with no embedded words fillers. 291
Appendix B.4 Embedded words with initial /l/. 291
Appendix B.5 Embedded words with initial /r/. 291
Appendix B.6 Embedded words with initial /w/ 292
Appendix C.1 The phonetic transcription task given to subjects one week prior to pre-tests. 292
Appendix C.2 A list of 12 illegal English consonant clusters given to subjects at the beginning of the treatment. 293

List of Figures

Figure 2.1 A simple model of human information processing system. Adapted from Carroll (1986:46) and Flowerdew and Miller (2005:23)..................................18
Figure 2.2 The old and new models of working memory. From Baddeley (2000: 418)..20
Figure 3.1 A spectrogram of the sentence *any bright young man who still thinks*. From Potter, Kopp and Kopp (1966)..64
Figure 3.2: Sonority Sequencing Generalization (from Broselow and Finer (1991: 37)..103
Figure 3.3 Weber and Cutler's (2006) Word Spotting Task results expressed as mean percentage reduction in RT and miss rate, for each listener group in each boundary condition (Common, English, German) compared with the *no boundary* condition (From Weber and Cutler 2006: 603)..113
Figure 5.1. Native English speakers' out of 5 mean ratings in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2)...174
Figure 5.2. Native English speakers' mean percentage of errors in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2)..174
Figure 5.3. Non-native control and non-native experimental groups' ratings in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2)..176
Figure 5.4. Non-native control and non-native experimental groups' mean percentage of errors in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2).......................................176
Figure 5.5. Mean ratings in the Non-word Rating Task of the three groups (native control, non-native control and non-native experimental) in the four phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability)..............178
Figure 5.6. Mean percentage of errors in the Non-word Rating Task for the three groups (native control, non-native control and non-native experimental) in the four phonotactic conditions (illegal in 1, illegal in 2, low probability and high probability). 178
Figure 5.7 Native speakers' mean RTs in ms measured from item onset in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability)..193
Figure 5.8 Native speakers' mean percentage of errors in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability)..193
Figure 5.9 Non-native control and non-native experimental groups mean RTs in ms measured from item onset in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability)..195
Figure 5.10 Non-native control and non-native experimental groups' mean percentage of errors in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in , low probability and high probability)..195
Figure 5.11 Mean RTs in ms in the LD Task for the three groups (native control, non-native control and non-native experimental) in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability)...............................196
Figure 5.12 Mean percentage of errors in the LD Task for the three groups (native control, non-native control and non-native experimental) in the four different phonotactic conditions)..196

Figure 5.13 Native speakers' mean RTs in ms measured from target word offset in the WS Task in four different boundary conditions. 210
Figure 5.14 Native speakers' mean percentage of errors in the WS Task in four different boundary conditions. 210
Figure 5.15 Mean RTs measured from target word offset for the two non-native groups in the Word Spotting Task in four different boundary conditions. 212
Figure 5.16 Mean percentage of errors of the two non-native groups in the Word Spotting Task in four different boundary conditions. 212
Figure 5.17 Mean RTs in ms measured from word offset in the Word Spotting Task for the three subject groups (native control, non-native control and non-native experimental) in four different boundary conditions. 213
Figure 5.18 mean percentage of errors in the Word Spotting Task for the three subject groups (native control, non-native control and non-native experimental) in four different boundary conditions. 214
Figure 5.19 Non-native control group's mean ratings in the NWR Task in four different phonotactic conditions in the pre-test and post-test. 221
Figure 5.20 Non-native control group's mean percentage of errors in the NWR Task in four different phonotactic conditions in the pre-test and post-test. 221
Figure 5.21 Native control group and non-native control group's pre-test and post-test ratings in the NWR Task in four different phonotactic conditions. 222
Figure 5.22 Native control group and non-native control group's pre-test and post-test errors in the NWR Task in four different phonotactic conditions. 223
Figure 5.23 Non-native experimental group's mean ratings in the NWR Task in four different phonotactic conditions in the pre-test and post-test. 224
Figure 5.24 Non-native experimental group's mean percentage of errors in the NWR Task in four different phonotactic conditions in the pre-test and post-test. 224
Figure 5.25 Native control and non-native experimental group's pre-test and post-test ratings in the NWR Task in four different phonotactic conditions. 225
Figure 5.26 Native control and non-native experimental group's pre-test and post-test errors in the NWR Task in four different phonotactic conditions. 226
Figure 5.27 Non-native control group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions. 231
Figure 5.28 Non-native control group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.231
Figure 5.29 Native control group and non-native control group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions. 232
Figure 5.30 Native control group and non-native control group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions. 232
Figure 5.31 Non-native experimental group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions. 234
Figure 5.32 Non-native experimental group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.234
Figure 5.33 Native control group and non-native experimental group's mean RTs in ms in the pre-test and post-test in four different phonotactic conditions. 235
Figure 5.34 Native control and non-native experimental group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions. 236

Figure 5.35 Non-native control group's mean RTs in ms in the Word Spotting Task in the pre-test and post-test in four different boundary conditions. 242
Figure 5.36 Non-native control group's mean percentage of errors in the Word Spotting Task in the pre-test and post-test in four different boundary conditions.....242
Figure 5.37 Native control group and non-native control group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different boundary conditions. 243
Figure 5.38 Native control group and non-native control group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different boundary conditions 244
Figure 5.39 Non-native experimental group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different phonotactic conditions 245
Figure 5.40 Non-native experimental group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different phonotactic conditions....245
Figure 5.41 Native control group and non-native experimental group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different boundary conditions. 247
Figure 5.42 Native control group and non-native experimental group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different boundary conditions. 247

List of Tables

Table 2.1 Four levels of language processing. Adapted from Carroll (1986: 108)...17
Table 3.1 Mean percentage missed targets (errors) and mean reaction times for correct detection (RT, in ms), measured from target-word offset in McQueen (1998).......101
Table 4.1 Non-native subjects' mean scores in the three criteria upon which the division process into control and experimental groups was based......................133
Table 4.2 Mean scores for the five statistics performed on the stimuli in four conditions..149
Table 5.1 Mean RTs in ms and mean percentage of errors for the 3 groups in phonotactic conditions...189
Table 5.2 Mean RTs in milliseconds measured from word offset and mean percentage of errors in the pre-test for native control, non-native control and non-native experimental groups in four different boundary conditions............................206

Chapter 1: Introduction

1.1 The problem

"Native speakers speak too fast" and "native speakers do not speak English". English as a foreign language (henceforth EFL) teachers are perhaps familiar with such comments from their students. EFL teachers expect to hear these comments when their students have the chance to listen to naturally produced speech in authentic teaching materials, through the media or even when they seize the few opportunities of communicating with 'inconsiderate' (i.e. in terms of the language level of the non-native interlocutor) native speakers. Students' complaints may have a real cause. As some researchers have noted (e.g. Brown 1990; Rosa 2002; Stanley 1978) the English language EFL students are taught and hear in the classroom is not representative of naturally produced English. EFL teachers often tend to speak slowly and articulate clearly (i.e. using, a register very much like foreigner talk, which is characterized by fewer connected speech phenomena (see Tarone 1980: 423), as discussed below, and trying to pronounce the citation form of words).[1] Furthermore, EFL listening texts are typically graded in the language they use (e.g. slow speech rate), giving the learners a false sense of ease in listening to EFL.

Unlike the slowly and carefully produced speech EFL learners are familiar with in the classroom, in naturally produced connected speech words are typically produced rapidly with a speech rate between 230-280 syllable per minute, i.e. 3.8-4.7 syllable per second considered average (Tauroza and Allison 1990). As a result, words in

---

[1] Let alone when the EFL teacher is non-native speaker whose mere pronunciation is non-target-like.

connected speech can undergo considerable modification to minimize the effort made to produce speech.

1.1.1 Connected speech

The phonological phenomena words undergo in connected speech cause their pronunciation to deviate from the citation forms which EFL learners might be used to. This therefore poses a problem for inexperienced EFL learners. Gimson (2001: 287-95) and Geigerich (1992: 249-90) discuss some of these phonological phenomena. Here I will describe three common phenomena which include reduction, elision and assimilation.

In connected speech segments are not pronounced as fully as they are in the citation form of words. Full vowels, for example, can be reduced. Such vowels in the citation forms of function words (e.g. 'of') and unstressed syllables in connected speech are usually reduced to a schwa[2] and sometimes short vowels and schwas are elided (i.e. deleted) completely in connected speech as in (1) where (a) is the citation form and (b) the pronunciation in connected speech.

1) I went to the federal police
a- /aɪ wɛnt tu ðə fɛdərəl pəli:s/
b- [əwɛntəðəfɛdrlpli:s]

The same forms of modification apply to consonant segments in connected speech. Like vowels they can sometimes be elided in connected speech as in (2).

2) I took him to the library
a- /aɪ tuk hɪm tu ðə laɪbrəri/
b- [ətukɪmtəðəlaɪbri]

---

[2] Or other central vowel.

Not only can consonant segments be elided, but they can also assimilate (usually in the place of articulation) to the following segment, be it in the same word or across a word boundary as in (3). In this sense, assimilation in English is often called "anticipatory" (Geigerich (1992: 288).

3) good card with ten points.
a- /gʊd kard wɪð ten pɔɪnts/
b- [gʊg kard wɪð tem pɔɪnts]

Sometimes, words in connected speech can undergo a combination of elision and assimilation processes as in (4) where as shown in (b) the /d/ in *and* was deleted then /n/ assimilated with the following /b/ and is realised as [m].

4) bed and breakfast
a- /bed ænd brekfæst/
b-[bedəmbrekfəst]

These connected speech processes cause listening difficulty for inexperienced second language (henceforth L2) listeners (Henrichsen 1984).

The current research is of the view that a great deal of L2 learners' listening problems are the result of lack of automaticity in the L2 bottom-up listening skills. Therefore, if we are to investigate the effectiveness of a bottom-up approach in L2 listening teaching we need to target a skill in instruction and then directly test the automaticity of this skill using a well-suited measure. The skill that is the target of instruction in the current study (i.e. lexical segmentation) is discussed in the next section.

1.1.2 Lexical segmentation

Determining word boundaries may be an important contributor to listening comprehension insofar as it aids the word recognition process. As native speakers of our first language (henceforth L1), we may not realize how complicated the task of lexical segmentation (i.e. determining word boundaries) is. However, when we listen to a foreign language we realize that this is indeed the case. It soon becomes clear that finding word boundaries is difficult (Field 2003). Unlike the case with written sentences where word boundaries are clearly marked by white spaces, in connected speech the only similarly reliable cue for a word boundary are the pauses the speaker may make every five or six words. Otherwise, the rest of speech is a continuous flow of sound, or in Brown's (1990: 2) words an "acoustic blur" with no clear boundaries between words. Consequently, non-native listeners may not only be unable to recognize new words but also known words can go unrecognized (Goh 2000).

Psycholinguistic research has shown that one way native speakers (including L1-learning infants) manage to overcome the lexical segmentation problem is by exploiting segmentation cues in the signal indicating word boundaries. These include allophonic cues (Nakatani and Dukes 1977; Jusczyk et al. 1999a), prosodic features (Cutler and Norris, 1988; Jusczyk et al., 1999b) and phonotactic rules (Mattys and Jusczyk 2001; McQueen, 1998). One problem for L2 listeners with these cues is that they are language specific. In other words, different languages have different variations of these cues. For example, whereas the phonemes /r/ and /l/ are context-sensitive and therefore have different allophonic variations in English (Church 1987; Nakatani and Dukes 1977), Japanese speakers do not have a phonemic contrast at all between these two phonemes.

Phonotactic constraints are also language-specific. Whereas clusters like [ps], [pn] and [pt] can be word-initial in German, French and Greek respectively, none of these clusters can be word-initial in English (Davenport and Hannahs 1998: 105).

Language specificity of segmentation cues is likely to cause problems for L2 listeners. It appears that long exposure to the L2 may not be sufficient to enable listeners to use L2-specific segmentation cues when listening to the L2. Indeed it has been shown that L2 listeners do not use the L2-specific prosodic segmentation strategies when listening to the L2 (Cutler et al. 1992). On the other hand, it was shown that highly advanced EFL learners use the EFL-specific phonotactic constraints by assuming word boundaries between illegal English sequences (e.g. /tl/) when listening to English (Weber and Cutler 2006). However, these learners were also using their uninformative L1 phonotactic constraints inappropriately (ibid). Therefore, while a German speaker might rightly assume a word boundary between the two phonemes in /tw/ when listening to German, s/he would missegment if s/he still assumed a boundary when listening to English and therefore might mishear '*be twins*', for example, as '*beat wins*'.

In the literature, phonotactics has not only been treated as categorical (i.e. legal vs. illegal) but also as probabilistic. Probabilistic phonotactics refers to "the relative frequencies of segments and sequences of segments occurring in syllables and words" (Vitevitch and Luce 1999: 375). It has been shown, using different tasks (Bailey and Hahn 2001; Vitevitch, Luce, Charles-Luce and Kemmer 1997; Vitevitch and Luce 1998; 1999; Vitevitch and Luce 2005) that phonotactic probability has varying effects on adults' speech processing of an L1. To my knowledge, no study has attempted to examine the effect of probabilistic phonotactics on the processing of EFL, yet this would show what role phonotactic probability can play in the acquisition of EFL.

The present study endeavors to provide insight into the role of phonotactics in the processing and lexical segmentation of EFL. More specifically, it aims to investigate if low-intermediate Arabic speaking EFL learners have acquired sensitivity to the phonotactic constraints and probability of English, and if they use the phonotactic constraints of their L1 (Arabic) or their FL (English) when segmenting connected speech in English (i.e. do they, like English native speakers, assume the presence of a word boundary upon hearing an illegal English cluster (e.g. /dl/) in a sequence such as *bad lady*?). More importantly, the present study also uniquely examines the effect of teaching English phonotactic constraints on the lexical segmentation ability of EFL learners in running speech in English.

1.2 Phonotactics of English and Arabic

As was stated, most segmentation cues including phonotactic constraints are language-specific. There are different types of proposed phonotactic constraints (e.g. Brent and Cartwright 1996): the "vowel constraint" which states that each syllable should contain a vowel; the positional constraint governing which sounds can appear in certain onset or coda positions in a syllable, and the co-occurrence or the sequencing constraint which determines which sounds can appear adjacent in the onset or the coda of the syllable. The first two constraints are discussed further in Chapter 3. The identification of word boundaries is especially aided by the third constraint, the sequencing constraint, which governs which sound classes can appear adjacent to each other within a syllable or word and deserves more attention here. In English, for instance, a labial segment (e.g. /p/, /v/) cannot be followed by /w/ in an onset or in a coda (Davenport and Hannahs, 1998: 147). Similarly, clusters such as /pr/ and /br/, although completely legal in onset position, (e.g. prime and bread), are illegal in coda position. The opposite is true for /nt/.

Important to note here is that Arabic has a diglossic situation where specific constraints vary. Arabic speakers typically learn their native Arabic dialect at home and then are taught standard Arabic (Henceforth SA) at school. SA is nowadays the native language of only a small number of Arabs in some parts of some Arab countries where SA is spoken at home.[3] However, it is learnt from the early stages in school. Additionally, Modern SA, a variety of Arabic that follows most of SA phonological rules, is used in formal lectures and most of the radio and TV programs. Therefore, the possibility that our Arabic speakers EFL have acquired sensitivity to SA phonotactic constraints besides those of their own native dialect can not be ruled out, a possibility that was taken into consideration when designing the tasks of the current study.

Differences include the fact that SA does not allow bi-consonantal clusters in syllable onsets. However, Arabic dialects including Qassimi Arabic (henceforth QA), the native dialect of our EFL subjects, may allow some. Procedures discussed in Chapter 4 were used to determine legal onset clusters in QA.

1.3 Context and significance of the study

The present study was conducted in the central region of Saudi Arabia. In Saudi Arabia, English is taught as a foreign (vs. second) language. In the present thesis, I

---

[3] In an attempt to preserve (M)SA, some Arabic speakers use (M)SA instead of their native Arabic dialect to communicate with their children at home. Therefore, these children usually learn (M)SA as their native language.

use the term EFL to refer to the setting where the target language (i.e. English) is not the medium of communication in the community. On the other hand, English as a second language (henceforth ESL) is used when referring to settings where English is the medium of communication. Although EFL and ESL might be essentially similar in terms of the learning process involved, ESL learners often have the advantage of being exposed to more naturalistic input from native speakers than EFL learners.

Most EFL learners, including those in the current study, are unlikely to be exposed to much naturalistic (i.e. from native speakers) input.[4] Explicit teaching may help them to make the best out of the little input they receive by directing their attention to the presence of relevant phonotactic cues. The current study therefore examines the effectiveness of this kind of awareness raising. Is the resultant knowledge applied to lexical segmentation? And if it is, how automated does this process become compared to that of native speakers? Specifically, the present study aims to investigate four research questions. These are as follows:

(1) Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English and Arabic?
(2) Are Arabic speaking EFL learners sensitive to the phonotactic probability of English?
(3) Do Arabic speaking EFL learners use the phonotactic constraints of English and Arabic in lexical segmentation of running speech in English?
(4) Can explicit teaching of English phonotactic constraints help improve Arabic speaking EFL learners' ability in lexical segmentation of English?

---

[4] Most English language teachers, including the lectures in the university where the present study was conducted, are non-native English speakers.

The present study is significant for a number of reasons. First, only a small number of studies have attempted to examine the role of phonotactic constraints in the segmentation of an L1, Research investigating the role of phonotactic constraints of the L1 and L2 in the segmentation of an L2 is scarcer still. Notably, this study includes subjects who are native speakers of Arabic, a language that is relatively understudied. In addition, it aims to provide insight into an area of research that to my knowledge has not been approached before, namely the role that English probabilistic phonotactics play in the processing of EFL. Importantly, it also endeavors to study the viability of introducing a new teaching method for L2 listening, that is, teaching phonotactic constraints.

Unlike most research and teaching methods in L2 listening which (unjustifiably as we will see) concentrate on top-down processing and listening strategies, the present study, as stated above is of the view that a great deal of L2 listeners' problems are the result of their lack of automaticity in bottom-up skills, most notably their lexical segmentation ability and hence students' discouraged comments discussed above. The role of the classroom in helping learners automate bottom-up listening skills is of paramount importance especially in EFL contexts given the lack of naturalistic input and consequently learners' inability to rely solely on natural acquisition of these skills.

It has been noted, however, (e.g. Doughty 2003) that studies investigating the effect of explicit teaching on L2 acquisition widely suffer from the defect of using measures that are metalinguistic in nature. Studies investigating the role of explicit teaching in improving learners' bottom-up listening skills have another disadvantage. That is, they have typically used global measures that fail to pinpoint the source of post-instruction improvement. To overcome this flaw, the current study used three different measures. First, it used a measure (Non-word Rating task) where subjects' metalinguistic knowledge can be used. In this task, subjects had to give their subjective judgments of the English-likeness of non-word items starting with illegal consonant clusters in Arabic and in English, non-words with low phonotactic probability and non-words with high phonotactic probability. Another on-line measure (Lexical Decision task) where the use of metalinguistic knowledge was unlikely was used to test subjects' repeats sensitivity to the same non-word items presented in the Non-word Rating task. Finally, the Word Spotting task was used to directly measure the targeted skill (i.e. lexical segmentation), and therefore was better able to pinpoint the source of improvement. In addition, unlike some empirical studies in L2 listening which used the difference between pre-test and post-test scores as the only measure of improvement, the current study also used a native control group so that non-native subjects' scores could be compared to ultimate native speaker performance.

1.4 Overview of the thesis

The main objective of the present study is to investigate the effect of teaching EFL phonotactic constraints on one of the most important skills (i.e. lexical segmentation) involved in EFL listening comprehension. For that purpose, Chapter 2 presents a detailed account of listening comprehension including definition of the term and cognitive processes involved (Section 2.2) and the role of memory in listening (Section 2.2.1). Types of processing in terms of their automaticity (i.e. automatic vs. controlled) and direction (i.e. top-down vs. bottom-up) which are used in listening comprehension are detailed in Sections 2.2.2 and 2.2.3, respectively. In Section 2.3, research in L2 listening instruction with the two widely adopted approaches (top-down and bottom-up) distinguished is reviewed.

Chapter 3 is concerned with the targeted skill (i.e. lexical segmentation). Different segmentation models (serendipitous vs. explicit) are reviewed in Sections 3.2 and 3.3, respectively. In Section 3.3, Explicit Segmentation Models (i.e. those that propose that cues in the signal aid lexical segmentation) are supported and three cues (allophonic, prosodic and phonotactic) under this category are discussed in Sections 3.3.1, 3.3.2 and 3.3.3, respectively. Research on the use of these cues by L1-learning infants and adult native speakers will also be reviewed. Section 3.3.4 sets the context of the present study by highlighting the processes that underlie the learning of the phonotactic constraints which are the focus of the current research. In Section 3.3.5 questions will be raised and answers attempted regarding the language specificity of these cues. Section 3.4 discusses how instruction can help in the acquisition of phonotactic constraints and their use in lexical segmentation. Section 3.5 highlights another form of phonotactics (i.e. probabilistic phonotactics) investigated in the present study. Questions that can be answered by investigating the effect of EFL phonotactic probability on the processing of EFL are also discussed.

Chapter 4 provides a detailed description of the methodology utilized in the present thesis. Sections 4.2 and 4.3 provide an account of the subjects and study design. The three tasks used (i.e. Non-word Rating, Lexical Decision, and Word Spotting) are detailed in Section 4.4, where the instructional treatment and relevant materials, procedures, research questions and hypotheses are also presented.

Chapter 5 reports on the results obtained from each task. Pre-test results are presented and discussed in Section 5.2, and post-test results are presented and discussed in Section 5.3. How these results provide answers to our research questions is summarized in Section 5.4.

Major findings are recapitulated in the last chapter (Chapter 6). Limitations (Section 6.2), pedagogical implications (Section 6.3), suggestions for further research (section 6.4) and general conclusions (Section 6.5) are also discussed. All materials used in the thesis appear in the appendix.

Chapter 2: L2 Listening Comprehension: Auditory Processing and Teaching

2.1 Introduction

Listening comprehension is a complex process that takes place through the integration and interaction of cues from different levels (e.g. phonetic, lexical) and sources (e.g. contextual and background knowledge) (Lynch 2002; Oxford 1993; Rost 2001). It is a skill that is of paramount importance in language acquisition (henceforth LA). This can be seen in L1 acquisition when children having congenital hearing problems may end up without linguistic competence despite their articulatory system being intact. The importance of listening in LA can also be found in empirical research which shows that native adults spend most of the time during communication listening rather than on any of the other language skills (i.e. speaking, reading, writing) (Rankin 1930 reported in Feyten 1991).

In L2A, listening is hardly less important. Most theories of L2A point out the importance of aural input for acquisition of linguistic competence (McLaughlin 1987).[5] However, not all input can become intake (Corder 1967). Some researchers suggest it is attended to or consciously noticed input that can become intake[6] (Oxford 1993; Schmidt 1990, 1995) while others suggest it is comprehended input (Krashen 1982, 1985, 1994). The amount of input that is converted to intake is critically dependent on a number of factors amongst which is the way it is processed (Sharwood-Smith 1986) and the 'direction' of processing (i.e. bottom-up or top-down). Knowledge of these aspects is really important for the teaching of L2 listening

---

[5] See Carroll (2001) for more discussion of this issue.

[6] The discussion of input vs. intake will be presented in Section 2.2.2.

as it can help us determine our methodology for teaching it. In other words, if our aim in teaching L2 listening is only to help learners comprehend the aural message, our methodology will be different from that if our aim was to help learners acquire the L2.

Although acquiring the L2 is, of course, the paramount aim for teaching it, the majority of research in L2 listening seems to have overlooked this point. In other words, research in L2 listening training seems to over-concentrate on developing learners' top-down and compensatory strategies at the expense of automatizing any bottom-up skills which would help them acquire the L2. Research, as will be discussed, for example has suggested developing learners' listening ability by helping them use background and contextual information, and non-verbal cues including pictorial aids, lip-reading and body language. This is a practice which I will argue below is unjustified and misses the point of L2 listening.

On the other hand, the relatively little research which has tried to investigate the effectiveness of adopting a bottom-up approach to teaching listening does not provide robust evidence regarding the efficacy of bottom-up teaching. The reason behind this, as will shortly be discussed, lies in the data collection measures used. These measures are so general that they fail to pinpoint the source of improvement after teaching. The current thesis is of the view that a great deal of L2 learners' listening problems is the result of lack of automaticity in bottom-up skills. Therefore, it is assumed that a good way of testing the effectiveness of a bottom-up approach is by measuring its effect in enhancing the automaticity of learners' performance on their perception of the L2.

This chapter aims to discuss the psycholinguistic processes involved in listening comprehension and review research in L2 listening training. In Section 2.2, it begins by providing some definitions of listening comprehension and gives an overview of the cognitive processes involved in listening. Section 2.2.1 presents ideas on different types of memory and discusses their important role in listening comprehension. The types of processing in terms of quality (*automatic* vs. *controlled*) and direction (*top-down* vs. *bottom-up*) in listening comprehension are detailed in Sections 2.2.2 and 2.2.3. Section 2.3 reviews research in L2 listening instruction on two widely adopted approaches, namely top-down and bottom-up. In Section 2.3.1 research favouring the top-down approach is reviewed. It will be argued that the current over-reliance in applied linguistics and English language teaching on the top-down approach is based on unwarranted claims and misses the point of L2 listening teaching. Section 2.3.2 reviews the scarce research adopting a *bottom-up* approach to teaching listening. Criticising the latter will mainly be concerned with the assessment measures adopted. Based on this criticism, a methodology for the teaching and assessment of listening comprehension ability at the bottom-up level is proposed, a methodology which I used in the current thesis. Section 2.4 provides a summary for the points discussed in this chapter.

2.2 Listening comprehension

A search of the literature on L2 listening for a good definition, to my surprise, showed that most of the publications, books included, did not try to provide a definition of listening comprehension, although some of the publications actually included the term *listening comprehension* in their titles (e.g. Brown and Hilferty 1987; Graham 2006; Richards 1983). This apparent trend of avoiding providing a definition for listening comprehension is perhaps not because the definition of listening comprehension is taken for granted; rather it seems that defining listening comprehension is problematic when we look more closely. Glen (1989) analyzed a number of definitions for native language listening and concluded that a universally accepted definition does not seem to exist. Rost (2001) confirms this and notes that different fields have provided different definitions of listening according to how it relates to their theoretical interests. Rost, however, defines listening as "experiencing contextual effects" (p 3). Admittedly broad, this definition does not provide any insight into the actual processes involved in listening comprehension. One of the few definitions I found to be adequate is provided by Vandergrift (1999: 168). In language teaching, Vandergrift defines listening comprehension as

> a complex, active process in which the listener must discriminate between sounds, understand vocabulary and grammatical structures, interpret stress and intonation, retain what was gathered in all of the above, and interpret it within the immediate as well as the larger sociocultural context of the utterance.

Despite the lack of agreement on a definition of listening comprehension, authors seem to agree that listening comprehension involves different types of processing. These include neurological, linguistic, pragmatic and psycholinguistic (Rost 2001). In cognitive psychology, Anderson (1983, 1995) differentiates three interrelated stages through which comprehension takes place. These are perception, parsing, and utilisation. In the first stage, perception, recognition and encoding of speech sounds takes place. This stage also includes word recognition after combining phonetic sounds into words. In the second stage, parsing, words recognised are assigned structure and are also used to construct meaningful mental representations. In the last stage, utilization, listeners may use the mental representation of the utterance's meaning and act accordingly.

Similar to Anderson, Carroll (1986) identifies four levels of processing the language in auditory or visual form. These are shown in table 2.1.

Table 2.1 Four levels of language processing. Adapted from Carroll (1986: 108).

Level	**Definition**	**Example**
Perceptual	Identify speech sounds	
Lexical	Retrieve the lexical +representation of a word from memory and integrate it into the ongoing context.	*firefighters*
Sentential	Determine the syntactic structure of the sentence as it is processed and store the gist of it	*Since she could not jump, the firefighters had to get her.*
Discourse	Identify the context preceding and following the a sentence and integrate the sentence representation with that context	*As the fire engulfed the opera house and drove both singers and customers to panic, there was a rush to the doorways and the soprano twisted her leg. Since she could not jump, the firefighters had to get her.*

The types and levels of processing involved in listening comprehension require a storage space from which meanings of words and relevant contextual information can be retrieved and also where new information can be kept while being processed. Different types of memory have been argued to be involved in the listening comprehension process. These are discussed in the following section.

2.2.1 Memory and listening comprehension

It is widely accepted that human memory is subdivided into three units: sensory memory or stores, short–term or working memory and long-term or permanent memory (Anderson 1995; Baddeley 1986, 1999; Carroll 1986; Gathercole and Baddeley 1993). The relationship and interaction of these three types and their role in information processing is depicted in Figure 2.1.

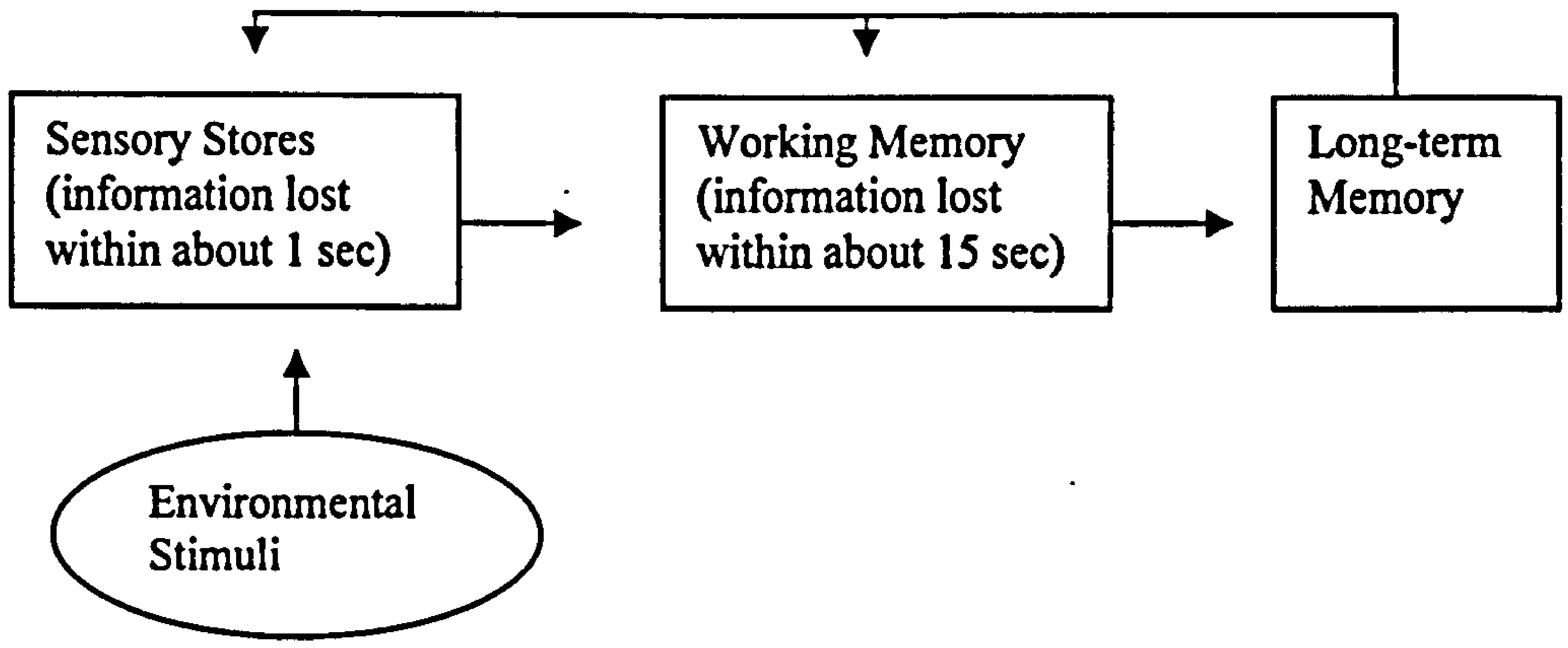

Figure 2.1 A simple model of human information processing system. Adapted from Carroll (1986:46) and Flowerdew and Miller (2005: 23).

Sensory stores are the unit that first receives the signal from the environment. It is divided into two types of stores: iconic memory, which is responsible for storing visual stimuli, and echoic memory which is responsible for storing auditory stimuli (Baddeley 1999). Auditory information does not stay long in the echoic memory (for about one second; Loftus and Loftus 1976). Depending on different factors, the stimuli can be passed to the working memory or just get lost. Processing of the attended stimuli starts in the working memory, which is limited in capacity (Anderson 1995; Carroll 1986). The capacity of the working memory is determined by the memory span which is measure by "the number of elements one can immediately repeat back" (Anderson 1995: 171), typically about seven (Baddeley 2000). After processing, the information may be transferred to long-term memory where it is stored for a longer time. Rehearsal of the message has been argued to play a role in determining whether the information can be passed to long-term memory, where meaning of the aural message rather than the exact form is stored, or whether it is lost (Clark and Clark 1977; Craik and Lockhart 1972; Rundus 1971). As the arrows in Figure 2.1 show, long-term memory does not just act as a permanent store; rather information in *long-term memory* can influence processing in the other memory units. As will be discussed in Chapter 3, our knowledge of the statistical probabilities of phonemes in words, for example, will result in a facilitative effect for repeating not only real words but also those non-words made up of frequently occurring phonemes (see for example Vitevitch et al. 1997).

Working memory is perhaps the most important unit in speech comprehension. It has been suggested that it is the place where sounds are stored long enough to be patterned into their appropriate syntactic units and subsequently interpreted semantically (Call 1985; Clark and Clark 1977). Knowledge of the way working memory functions and its components is helpful in gaining insight into its role in language comprehension; see Baddeley (1999) and Gathercole and Baddeley (1993) for a detailed discussion of the contribution of working memory to different aspects of language processing.

Baddeley (1999: 19) defines working memory as "the temporary storage of material necessary for performing a range of complex tasks such as comprehension, reasoning, and long-term learning". During the last four decades, a model of working memory has been developed, its theoretical accounts investigated and refined (Baddeley and Hitch 1974; Baddeley 1986, 1990, 2000, 2003). Baddeley and Hitch, for example, suggested that working memory has three components: the *central executive*, which is the main component, and which is supplemented by two other "slave systems": the *phonological loop,* which is responsible for maintaining and processing of verbal information, and the visuo-spatial sketchpad, which is responsible for the processing of visual information and therefore will not be further discussed here. Baddeley

(2000) added a fourth component, the *episodic buffer*, to the model. Figure 2.2 shows the old working memory model (Baddeley and Hitch 1974) in (a) and Baddeley's (2000) new four–component model in (b).

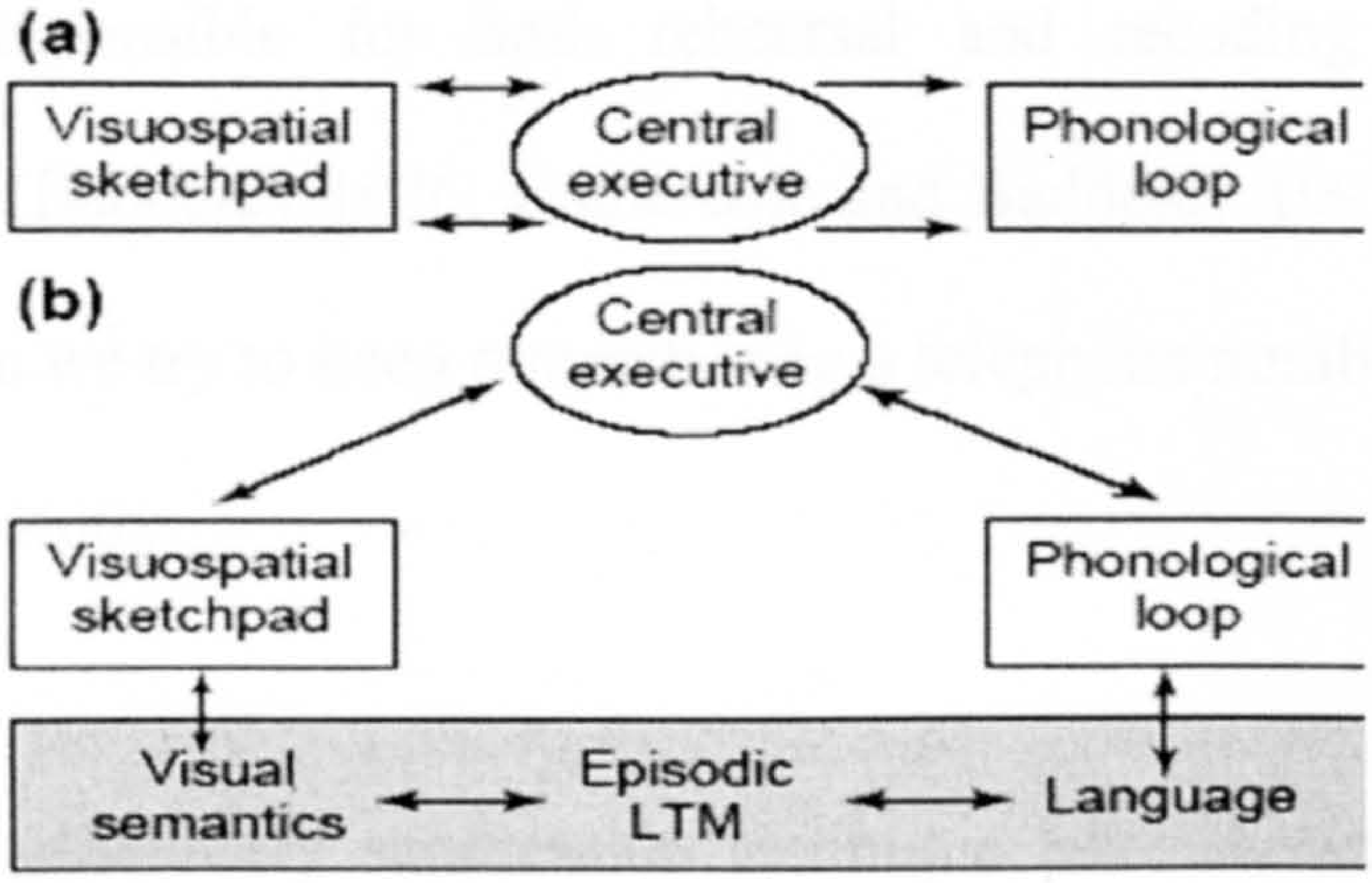

Figure 2.2 The old and new models of working memory. From Baddeley (2000: 418).

According to Baddeley, the episodic buffer acts as

> a limited –capacity temporary storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive, which is capable of retrieving information from the store in the form of conscious awareness, of reflecting on that information and,where necessary manipulating and modifying it (p. 421).

As its name suggests the central executive is the most important component of working memory as it acts as a supervisory control system. Its functions include controlling the transmission of information between other parts of the cognitive system, allocating appropriate input to the phonological loop and the sketchpad systems and retrieving information from long term memory (Gathercole and Baddeley 1993).

The phonological loop might be of most relevance to phonological learning as it has been shown to play a role in language acquisition, particularly vocabulary acquisition

(Gathercole and Baddeley 1989; Service 1992). The phonological loop is assumed to consist of a temporary phonological store responsible for storing verbal information for which memory traces may decay within a few seconds unless rehearsed. Rehearsal is controlled by the second component of the phonological loop which is the subvocal control process responsible for both rehearsal and recoding information into phonological form (Baddeley 1986; Gathercole and Baddeley 1993). An example of this process is when we try to keep remembering a telephone number by repeating the digits out loud or silently.

Studies using the articulatory suppression technique have shown that rehearsal of phonological information in the phonological loop is important for keeping phonological information longer in working memory. In an articulatory suppression study, subjects are visually or aurally presented with a memory list (words or numbers), and at the same time are asked to repeat an irrelevant word (e.g. *the*). It has been shown that using this technique, subjects' memory span decreases, suggesting that they are denied the benefit of rehearsing in the phonological loop (Estes 1973; Peterson and Johnson 1971). Using the same technique, Papagno, Valentine, and Baddeley (1991) investigated if rehearsal in the phonological loop plays a role in the learning of foreign language words. In one of their experiments, they presented adult Italian subjects with two lists of eight pairs of items. In one list of items, each pair contained two familiar Italian words. The other list contained a familiar Italian word and an unfamiliar Russian word. In one condition, stimuli items were presented auditorily and articulatory suppression was used. Papagno et al. found that in this condition recall of foreign words was worse than recall of familiar Italian ones. These results were taken as evidence that adult subjects' ability to recall foreign language words was impaired by articulatory suppression, suggesting an important role for the phonological loop in L2 vocabulary learning.

As was discussed, working memory seems to play an important role in language comprehension. It is the place where perceptual input is kept for linguistic processing to take place. Not only that but it - particularly the central executive component - can seek help by retrieving information, such as schemata or background knowledge from *long term memory* to help in the parsing and interpretation of input (Buck 2001). However, one of the features of working memory is that it is limited in capacity and processes compete for limited resources (Anderson 1995; Carroll 1986; Gathercole and Baddeley 1993). Native speakers can compensate for this limit in capacity by performing some of the processes (e.g. word recognition) automatically (Carroll 1986; Vandergrift 2004) leaving adequate resources for other attention-demanding processes as will be discussed in the next section. Therefore, attentional resources can be freed for processing input fairly quickly. However, because of inadequate perceptual processing abilities, L2 learners may not be able to process the input (particularly recognising words in connected speech) automatically and tend to process the input in a controlled way instead (Flowerdew and Miller 2005, O'Malley, Chamot and Küpper 1989). Crucially, it has been shown that memory span for foreign language input might be even more limited than for native language input (Call 1985). Therefore, because of the limitation of working memory, L2 learners often find themselves lagging behind when processing connected speech, and as a result their comprehension may break down (Vandergrift 2004). The role of automatic and controlled processing in language comprehension is discussed in the next section.

2.2.2 Controlled vs. automatic processing.

The implicit knowledge which native speakers have of their L1 has often been associated with automatic performance. Paradis (1994), for example, claims that implicit knowledge is that which is used automatically. Similar claims are made by other researchers in L2 acquisition (e.g. Bialystok 1994; DeKeyser 2003; Ellis 1994a) who either explicitly or implicitly suggest that automaticity and implicit knowledge are two related notions (but see Ellis 1994b for a different view).

In addition, the issue of automaticity in language acquisition is rooted in the modular view of languages. In his theory *The Modularity of Mind,* Fodor (1983) claims that modular processes operate automatically, unconsciously and without control.[7] To explain his point he refers to the way native speakers listen to their L1. Although they can choose not to listen to their L1 by stopping their ears for example, once they listen to it they can not choose not to hear it. Similarly, we can not choose not to see things when we open our eyes as both faculties (i.e. processing the language and seeing) are modular (see also Schwartz 1993).

Researchers holding a general cognitive view of learning an L2 (e.g. McLaughlin 1987; McLaughlin, Rossman, and McLeod 1983) have a different position than Fodor's (see also Schwartz 1993) on what it means to learn a language and how automaticity is achieved. For them, the fluent (i.e. automatic) performance of advanced L2 learners is not the result of the control of a separate module of the mind. Rather, it is because "to learn an L2 is to learn a skill" (McLaughlin 1987: 133) and

---

[7] Modules have many other characteristics. See Fodor (1983) for a discussion.

that skill learning, including L2 learning, takes "the automatization of component sub-skills" (ibid).

Lack of automaticity of listening skills is a major factor in unsuccessful L2/FL listening. Therefore, helping learners automise their lower-level (bottom-up) skills is a good way of improving their listening ability because more attention will be available for higher-level (e.g. semantic, pragmatic) processing (Rost 2001; Segalowitz 2003).

In the information-processing approach of skill acquisition (Anderson 1982; 1995) a skill is acquired through integrating pieces of information into complex procedures.[8] In the early stages of skill acquisition only controlled processing is used. Here, the L2 learner consciously controls the process of coordinating pieces of information, which results in these processes occupying a large part of short-term (working) memory. With time and practice subcomponents that are required for carrying out a skill are gradually routinized and are eventually subsumed under one procedure which is stored in long term memory and which, once triggered, can automatically control all the sub-processes subsumed under it (Hulstijn 1990). Thus, carrying out this skill is no longer attention-demanding.

This perspective on L2 learning assumes a transition from early stage controlled processing to a developed stage automatic processing which Segalowitz (2003) illustrates with the change in learning how to drive a car. At the first stages one tends to pay careful attention to each sub-skill of the driving process, where for instance, one may tend to look at the gear while shifting. One may also be easily interrupted,

---

[8] See McLaughlin and Heredia (1996) for a review of the assumptions regarding this approach.

when spoken to for example, and feel that driving is a very effortful task. Gradually, the need to pay attention decreases so that we need not look at the gear while shifting and we can talk to passengers while driving. At the early stage, when processing is controlled and more cognitively demanding (Ellis 1994b:85), conscious attention is necessary, but when the performance later becomes automatic through practice, attention is no longer necessary and can be devoted to other related skills that have not yet been automatized (McLaughlin, Rossman, and McLeod 1983; Segalowitz 2003; Shiffrin and Schneider 1977).

Another distinction relevant to our discussion of automaticity and how it is achieved (see the example above) is between declarative knowledge and procedural knowledge. Anderson (1983; 1995) claims that information is stored in long-term memory as either declarative knowledge or procedural knowledge. Declarative knowledge "refers to consciously held skill-relevant knowledge that is describable" (Segalowitz 2003: 395). Procedural knowledge on the other hand, cannot always be verbalised and "consists of processes and skills, or the things that we know "how" to do" (Chamot and O'Malley 1994: 376). Under this view the knowledge of an L2 grammatical rule is declarative, but the way native speakers apply this grammatical rule fluently and unconsciously is procedural knowledge. *The Atomic Component of Thought Theory* (Anderson and Lebiere 1998) assumes that procedural skills are automatic and brought about through a gradual progression from declarative knowledge to procedural knowledge through a process called proceduralisation.

This is in line with the position proposed by some researchers (e.g. Hulstijn 1990; Levelt 1978; Schmidt 1992) who claim that language acquisition entails the acquisition of procedural skills. In this framework as Hulstijn (1990) has pointed out language acquisition can be viewed as the result of the compilation of declarative knowledge through which procedural knowledge could be established. Language automaticity will increase as a result of more sub-processes being subsumed under one complex procedure.

Anderson (1983; 1995) proposed a three-stage sequence in which proceduralisation takes place. These three stages are clearly described by Chamot and O'Malley (1994: 379):

> In the first stage, the learner approaches the new skill with conscious attention to rules and performs the skill deliberately and, most likely, with many errors. In the second stage of skill acquisition, some of the initial errors are eliminated and the performance becomes some what more fluent, though still not automatic. In the third stage, the performance is fine-tuned so that it becomes virtually automatic. At this point, the skill is said to be proceduralized, or converted from controlled or explicit processing to automatic or implicit processing.

However, in normal circumstances native speakers' implicit procedural knowledge of their L1 characterized by automatic performance develops first (Crain 1993). It is only at school age, through schooling or other sources, that they normally start developing declarative explicit knowledge of their L1.[9] Therefore, it is clear that this approach does not best describe how L1A acquisition takes place. That implicit procedural knowledge can develop first might also be true for naturalistic L2 learners who might show some competence of the L2 structures without necessarily being able to deduce the rules governing their performance (see Vainikka and Young-Scholten 2007). Proceduralisation may however be representative of how some L2 structures

---

[9] See also (Anderson and Fincham 1994: 1322) reviewing evidence against the claim that declarative knowledge is a prerequisite to procedural knowledge.

are learned in instructed L2A/FLA. However, before making any suggestions for pedagogical practices we need to understand what it means to automatically process the language in actual performance and what role formal instruction can play in that.

Judging linguistic performance as automatic falls out from the criterion of automaticity. Recall that the transition from controlled processing to automatic processing in L2A was described as a transition from an early stage when attention is needed to a later one, after practice, when attention is no longer needed. Johnson (1996) defines automization (i.e. automaticity) as "the ability to get things right when no attention is available for getting them right" (p 137). This "no need" for attention allows automatic processing to outperform controlled processing and consequently results in special characteristics of automatic performance.

Two widely agreed characteristics of automatic performance are speed and parallelness (e.g. Levelt 1989; Schneider and Fisk 1983; Segalowitz and Segalowitz 1993; Shiffrin and Schneider 1984). 'Parallel' means that more than one automatic process can be performed simultaneously with no or little interference between them. Investigating automatic performance in this sense would require designing a methodology in which simultaneous tasks are performed. However, Johnson (1996) has suggested that using computer technology, putting subjects under time pressure, is an equally proper distraction technique.

Various views have been proposed as to how formal language instruction can help in automising skills. One position stresses the role of formal instruction in providing the learner with the opportunity to practice and so to automatise new rules (e.g. Anderson and Fincham 1994; Sharwood-Smith 1981). This position claims that practice can turn non-automatic explicit knowledge into automatic implicit knowledge.

There is also the view that instruction can also indirectly foster automisation because of the role it plays in L2 acquisition. This view emphasises the role that instruction plays in conscious "noticing", which has been claimed to be necessary in converting input into intake as in the *Noticing Hypothesis* (Robinson 1995; Schmidt 1990, 1994, 1995. 2001). A distinction has been made between when input helps in the interpretation of an utterance and when it helps in constructing interlanguage (henceforth IL) grammar (Corder 1967). For the latter, the term *intake* is defined as "an abstract entity of learner language that has been fully or partially processed by learners, and fully or partially assimilated into their developing system" (Kumaravadivelu 1994: 37), whereas the former (i.e. input) is generally referred to as "what is available for going in" (Gass and Madden 1985: 3). Schmidt maintains that input must be consciously noticed by the learner in order to become intake. The motivation behind this proposal as Schmidt (1990) has argued is that, whereas L1 children are able to notice the L1 linguistic features while paying attention to the message, adult L2 learners have to notice these features in a more controlled way.

In addition, Schmidt and Frota (1986) argue that learners' focus of attention and noticing of mismatches between the input and their output determines whether or not they will progress in their acquisition. A similar claim is made by Gass (1990, 1991). Unlike Krashen (1982, 1985, 1994) who claims that comprehensible input alone is sufficient, Gass argues against the sufficiency of comprehensible input in building IL grammars. She maintains that if L2 learners are to convert input into intake, not only do they have to comprehend the input but also they have to notice the mismatches between the input and their IL grammar. In either case, formal instruction can play the role of an "attention getting" device that can help learners notice these mismatches (Gass 1991). See also White (1987) for a detailed argument against Krashen's *Comprehensible Input Hypothesis.*

According to this model, the mere presentation of comprehensible input is necessary but insufficient in converting certain constructions into intake.[10] If these constructions are to convert into intake they should be noticed, either because they are very salient or by the use of different strategies amongst which formal instruction has a prominent position. Empirical support for this position is found in some studies which Long (1996) cites. Some of these studies show premature stabilization or failure to incorporate basic target language structures by adults with prolonged exposure to comprehensible input. Other studies show global benefit for instructed language learning over purely naturalistic acquisition. Therefore, in this position, formal instruction can facilitate the process of noticing by directing attention to some linguistic features that could otherwise go unnoticed. Once knowledge is acquired, linguistic performance based on these feature is expected to be automatic.

If EFL teachers' main goal is to help learners automise their listening skills (Field 1998; Rost 2001), teachers will need to understand which level (i.e. perceptual, lexical etc.) and direction (i.e. top-down or bottom-up) of processing is more beneficial to target. A discussion of the direction of processing in listening is presented in the next section.

---

[10] This model assumes that there is a critical period after which natural L2 acquisition (i.e. relying merely on naturalistic input) of some linguistic features is unlikely.

2.2.3 Top-down and bottom-up processing

In section 2.2, types and levels of listening comprehension processes were discussed. As Carroll (1986) stresses, the presence of levels of processing does not imply that comprehension takes place in a serial order. Rather, the speed with which listening takes place suggests that processes at different levels operate simultaneously, as discussed in the previous section, and influence each other (Field 1999). The direction of these processes has often been depicted, both in ELT and Psycholinguistics, in two widely used terms, namely *bottom-up* processing and *top-down* processing. *Bottom-up* processing refers to "decoding the sounds of a language into words, clauses, sentences, etc. and using one's knowledge of grammatical or syntactic rules to interpret meaning" (Norris 1995: 47) or "that part of the comprehension process in which the understanding of the "heard" language is worked out proceeding from sounds to words to grammatical relationships to lexical meanings" (Morley 2001: 74). When we process bottom-up we follow a fixed sequence similar to that provided by Vandergrift's (1999) definition of listening comprehension presented above. We listen to individual sounds, combine them into words, combine words into phrases and sentences and finally interpret these sentences based on our background and contextual knowledge of the situation where the sentence was said (Brown 1990). This view of speech processing, which was developed during the 1940s and 1950s (Brown 1990; Flowerdew and Miller 2005), assumes that we process speech in a linear and hierarchical fashion. Five of the most important bottom-up skills in listening comprehension listed by Richard (1983: 228) are:

- Ability to discriminate among the distinctive sounds of the target language.
- Ability to recognise the stress patterns of words..
- Ability to identify words in stressed and unstressed positions.

- Ability to recognise reduced forms of words.
- Ability to distinguish word boundaries.

However, studies in L1 listening and psycholinguistics have shown that *bottom-up* processing is not the only way we process speech. Knowledge at a higher level can influence knowledge at a lower level. Knowledge of words, for example, could help improve speed and accuracy of recognising phonemes and hence involve what is known as the phoneme-restoration effect (Samuel 1981, 1991, 1996a; Warren 1970; Warren and Warren 1970). Warren (1970) aurally presented subjects with the sentence "the state governors met with their respective legislatures convening in their capital city". In the study, for example, the middle *s* in *legislatures* was replaced by a 120-millisecond cough. Only one subject out of twenty was able to detect that the phoneme was missing due to this effect.

Warren and Warren (1970) investigated if phonemes could be restored as a result of sentential context. They presented their subjects with sentences like the following where * denotes that the phoneme was replaced by non-speech:

a- It was that the *eel was on the axle.

b- It was that the *eel was on the shoe.

c- It was that the *eel was on the orange.

d- It was that the *eel was on the table.

Warren and Warren found that subjects reported hearing "wheel, heel, peel and meal", respectively, showing that phoneme restoration was affected by the context.

Many later studies have showed the same effect (see Samuel 1996b for a review). It seems therefore that speech can be processed top-down as well. However, because of inadequate knowledge of vocabulary, L2 listeners may not be as good in making use of word and sentential context in recognizing phonemes. That is because phoneme restoration has been found to correlate with other factors such as word frequency, neighbourhood characteristics (Samuel 1996b) which can not be matched with native speakers for FL/L2 learners. Therefore, FL/L2 listeners may need to rely more on *bottom-up* processing at least at lower levels.

In addition, in second language pedagogy, *top-down* has usually been used only as synonymous with background or contextual cues. Norris (1995: 47) for example defines top-down processing as "using background knowledge of the situation, context and topic to interpret meaning". When we do this we use our background and all types of contextual knowledge to predict what the speaker is likely to say in such situations. In this model (i.e. top-down processing), we do not need to listen to each feature to process the input; rather we monitor some parts of the acoustic message to confirm our expectations (Brown 1990). Brown (1994), Peterson (1991) and Richard (1983) outlined the micro-skills that are required in listening comprehension. Three of the skills at the top-down level are:

- Making inferences using real world knowledge.
- Getting the gist.
- Getting the meaning of the words from context.

Granted, listening does not seem to follow a fixed direction. Rather, listening is an interactive multi-directional process where different cues from different levels interact and are employed simultaneously. The nature of this interaction remains to be clearly identified.

L2 researchers have investigated the effects that these two types of processing training (i.e. top-down and bottom-up) have on L2 listening comprehension ability, although studies investigating the effect of the latter type are relatively scarce, as we will see. These studies are reviewed in the following section.

2.3 Teaching/training L2 listening

2.3.1 Introduction

Various approaches to teaching second language listening have been adopted, one of which is based on the premise that practice makes perfect. Following this approach, learners are given abundant in-class practice in listening and left to their own devices to acquire the listening skills (Ridgway 2000). The role of the teacher in such classrooms is to play tapes and check students' comprehension by asking them to answer written questions, to get the gist, or having them provide a summary. Consequently, classrooms are places where listening is in effect tested rather than taught (Sheerin 1987). Although such practice prevailed more than two decades ago, it is still predominant in some EFL contexts, including the Saudi context, as discussed in Chapter 1.

Another view is that listening classes can play a more active role by enhancing learners' listening skills. The debate, here is whether to target bottom-up processing or top-down processing.

One proposal is that using top-down cues including visual ones can enhance listeners' comprehension abilities (e.g. Ginther, 2002; Herron, et al. 1998; Jones and Plass 2002). Here the role of *top-down* processing is emphasised with an explicit call for teaching *top-down* cues and listening strategies. Consequently, L2 listening comprehension teaching has been dominated by top-down training, as noted by some researchers (e.g. Cauldwell 1996; Field 2003; Rost 2001; Wilson 2003), and developing *bottom-up* processing is ignored. Mendelsohn (1998), who is supportive of a top-down strategy-based approach to teaching listening, describes the change in practice in teaching L2 listening in the last half a century or so. He notes that

> There has been a shift from non-teaching in the audio-lingual period ("They'll pick it up by osmosis"), to haphazard listening to texts (many being readings of written language) followed by comprehension questions, to a "strategy-based approach" (Mendelsohn's term, 1994) in which students are taught strategies- that is , they are taught how to listen. (Mendelsohn 1998: 81)

Mendelsohn describes the strategy-based approach (described in detail below) as the method that teaches students how to listen. He seems to forget the bottom-up approach that if appropriately applied, in my view, it can not only teach L2 learners how to listen but also how to learn through listening.

The emphasis on a bottom-up method can be seen in books such as *Listening to Spoken English* (Brown 1990) which allots six chapters of a seven-chapter book to describing natural and connected speech phenomena and some phonetic as well as bottom-up cues (e.g. stress). Only parts of the last chapters discuss processing at the top-down level (e.g. learning to use contextual information).[11]

---

[11] However, this book is intended as a reference book rather than a student text.

However, the emphasis on a bottom-up approach to teaching listening by Brown and others does not seem to be reflected in EFL materials and classroom practices. As mentioned above, the fact that FL/L2 listening comprehension teaching has been and still is dominated by top-down training is widely recognised (Cauldwell 1996; Field 2003; Norris 1995; Rost 2001; Wilson 2003). Perhaps, one of the reasons behind this is that the relatively little research which tried to investigate the effectiveness of adopting a bottom-up approach to teaching listening could not provide reliable conclusions regarding the effect of bottom-up teaching.[12] The reason, as will be discussed below, lies mainly in the measures used. As already mentioned, these measures are so general that they fail to pinpoint the source of improvement after teaching.

The next section aims to review in detail how the effect of instruction on L2 learners' listening comprehension ability has been investigated. Studies in this domain can be provisionally divided into two categories. First, we have studies whose main concern, as in the present study, has been teaching L2 phonology and the effect of this on FL/L2 learners' listening abilities, in other words, taking a bottom-up approach to the teaching of FL/L2 listening. The main criticism of these studies will concern their measures. On the other hand, we have abundant research which has tried to investigate the effect of teaching top-down processing skills (e.g. compensatory strategies) on listening comprehension. Based on some assumptions discussed below, studies of the latter type have tried to improve listening comprehension by shifting the listener's focus from the real acoustic input to top-down cues. Assumptions and practices of the top-down approach will be critically examined.

---

[12] Akita (2005), for example, discussed below, investigated the effectiveness of providing ESL learners with perception training but, due to data collection method used, failed to pinpoint the source of improvement.

Before presenting the studies in these two categories we need to highlight a predominant method in FL/L2 listening teaching. Using this method which I call the "passive method" learners are given abundant practice in listening with ample exposure to comprehensible input and left to their own devises to acquire the listening skills (Krashen 1981, 1982, 1985, 1994, 1996; Ridgeway 2000). Within the framework of this approach, Krashen (1996) proposes what he calls *narrow listening* in which L2 listeners try to engage in conversation with native speakers or try to listen to authentic text at their own pace. Similarly, Ridgeway (2000) asserts that listening practice is the key to improving listening ability. Neither researcher, however, provides practical suggestions of how teachers can intervene to help student improve their top-down or bottom-up skills.

2.3.2 Teaching top-down skills

Some researchers examine what type of learning strategies and cues listeners use at the top-down level and how using and teaching these cues could help improve FL/L2 learners' listening comprehension ability. Relevant to top-down processing training is strategy instruction. Chamot (1995: 15) and O'Malley, Chamot and Küpper (1989: 422-23) differentiate three types of learning strategies: *Metacognitive* strategies, which plan, regulate and manage learning; *cognitive* strategies, which facilitate comprehension, such as making inferences based on background knowledge and guessing the meaning of words out of context; and *social* and *affective* (or socio-affective) strategies such as negotiating meaning and asking for clarification.

There are some problems with the concept of strategy training. First, there is the problem of definition (Ridgeway 2000) and whether strategies are conscious or unconscious. Second, strategies largely depend on the learner's characteristics, particularly whether s/he is a risk taker or a risk avoider (Field 1998), which raises the question of the practicality of teaching learners new cognitive strategies that they are not familiar with and how learners could utilise them.

The last problem concerns the classification of learning strategies. The classification provided by Chamot (1995) and O'Malley et al. (1989) is widely used in the literature. However, there is sometimes overlap between categories with no clear line between them (Rost 2001), particularly between metacognitive and cognitive strategies. For clarity I use the term 'metacognitive strategy' to refer to strategies which listeners use before and after the listening task to manage, plan and evaluate their learning (Vandergrift 1999). Metacognitve strategies can also include other strategies which learners use to learn the language, such as watching movies or listening to L2 materials or even trying to be in contact with native speakers. I use the term 'cognitive strategy' to refer to strategies which learners apply while listening to make the text more comprehensible. These include using world knowledge, visual cues, body movement and guessing meaning of unknown words from context. So, in a way cognitive strategies will be equated with compensatory strategies, a term used by Field (1998). It is training in strategies of the latter type that might shift the learners' focus from the real aural message.[13] This is a practice which I discuss critically below.

The main point in strategy training is that there are claimed to be certain effective strategies (e.g. guessing the meaning of words from context) that skilful learners use when listening. So, it is assumed that teaching learners how to use these strategies will

---

[13] See Thompson and Rubin (1996: 336), discussed below, for evidence that some of their subjects found visuals, for example, distracting.

help improve their listening comprehension ability (Goh 1997, 1998, 2000 2002; Lynch 1998; Mendelsohn 1994, 1995, 1998; O'Malley et al. 1989; O'Malley and Chamot 1990; Thompson and Rubin 1996; Vandergrift 1996, 1997, 1999). Mendelsohn describes the strategy-based approach and how it can be applied. He states that

> A strategy based approach is a methodology that is rooted in strategy instruction. It sees the objective as being to teach students how to listen. This is done, first, by making learners aware of how the language functions and second, by making them aware of the strategies they use-i.e., developing "metastrategic awareness". Then the task of the teacher becomes to instruct the learners in the use of additional strategies that will assist them in tackling the listening task (1995:134).

One of the compensatory strategies that L2 listeners seem to use is background knowledge. Long (1990) examined the role that background knowledge plays in the L2 listening comprehension of 188 L2 learners of Spanish. Subjects were asked to listen to different passages: one that was considered familiar and therefore allowing the use of background knowledge (rock music group U2), and another that was considered unfamiliar (a gold rush in Ecuador) and therefore did not provide subjects with as many top-down cues. Prior to listening to the two passages subjects were surveyed regarding their knowledge of this gold rush (in the first listening task) and rock groups (in the second) to help them activate their background knowledge of these two topics. Subjects were then presented with the listening tasks containing passages talking about the two topics and then asked to summarise the content of the passage in their L1 (English). They were also provided with a checklist containing distractors and asked to check only the statements that were included in the passage. Long found that learners' comprehension was better when listening to the familiar passage, suggesting that background knowledge facilitated the comprehension of the text.

The effect of advance organizers as top-down cues has also been examined. Advance organizers refer to the text-relevant information provided to listeners prior to the task. Herron, Hanley, and Cole (1995) compared the effects of two advance organizers for introducing 39 beginning English-speaking FL learners of French to video. The researchers had two conditions: a description-plus-pictures condition and a description-only condition. In the description-only condition the teacher read aloud a six-sentence summary of the major scenes in the video to be viewed, whereas in the description-plus-picture condition the teacher also showed a related picture while reading each summary sentence. Using comprehension questions as a measure Herron et al. (1995) found that a description-plus-pictures condition significantly improved listening comprehension over the description only advance organizer condition.

In another, similar study Herron, Cole, York and Linden (1998) examined the effect of two types of textual advance organizers. They had three beginning FL French groups: a declarative group, which listened to a six-sentence summary of the passage, an interrogative group, which listened to the same summary but in question form, and a control group, which was not provided with any summary. Herron et al. (1998) hypothesised first that the declarative and the interrogative groups would outperform the control group, and second that the interrogative group would outperform the declarative group as they assumed that questions would engage learners and concentrate their attention in the text more than the declarative summary would do. Comprehension questions which they asked learners in all groups after the task confirmed their first hypothesis. The two groups which were provided by advance organizers outperformed the control group. However, their second hypothesis was not confirmed as the interrogative group did not outperform the declarative group. Herron et al. attributed the latter finding to the fact that the questions given to the interrogative group did not match those asked in the final comprehension task, a mismatch which might have led subjects in the interrogative group to focus on irrelevant information. Also, subjects in the interrogative groups the authors reasoned may have focused too much on finding answers to the previewing questions and therefore missed other test-related information. However, attributing the better performance of the declarative and interrogative groups to advanced organizers per se should be treated with caution. The two groups may have outperformed the declarative group merely because for them advance organizers represented a somewhat first presentation of the information in the text which consequently helped them retain the information in the passage.

Visual cues have also been argued to benefit comprehension at the top-down level. Ridgeway (2000), for example, has stressed the importance that listening comprehension materials resemble real life listening situations, by using more video rather than audio materials. According to Ridgeway, this way, listeners can use visual cues like lip-reading and body language to aid comprehension, a suggestion that was previously made by Kellerman (1990). Actually, this is the type of top-down cues whose teaching might be counterproductive from an L2A perspective. These cues focus L2 listeners' attention away from the real perceptual input. Such practice may help learners *learn to listen* but certainly not *listen to learn* (the L2). A detailed argument against such an approach will be presented below.

Visual cues that are not authentic to the listening situation do not seem to help. Ginther (2002) compared the effect of content visuals (pictures that are related to the content of the conversation), which provide visual support of the text, and context visuals, which are pictures that are not related to the actual content of the text but rather set the scene for it. Ginther found that in the computerised TOEFL test, content visuals have only a little effect on comprehension whereas context visuals are actually counterproductive. A plausible explanation for this result is that visual cues that are not authentic to the listening task are processed independently and therefore draw the listeners' attention away from the information included in the aural message by consuming memory resources which would otherwise have been used to process the input (Vandergrift 2004).

Not only the help of inauthentic visual cues is limited but it also seems that its effect is dependent on the learner's proficiency level. In other words high proficiency level learners do not seem to benefit from visual cues. Mueller (1980) found that whereas less proficient subjects (English learners of German as an FL) benefited from pictorial aids while listening, more proficient students showed very little difference between listening with or without images. He suggested that that whereas pictures could help low-level students fill in the gaps that would otherwise exist in their prior knowledge, high-level students do not benefit from these aids because such students are "able to derive a context from the linguistic cues provided" (ibid: 340). This study provides indirect evidence that bottom-up-skills may be more important. In other words, top-down cues are only compensatory, and once bottom-up skills are mastered, as in the case of high-level learners, top-down cues may no longer be required.

Since some top-down cues were found to be helpful in listening comprehension as some of the studies discussed above showed, some researchers have reasoned that instructing learners in the use of these cues might improve their listening comprehension ability. O'Malley (1987) investigated the effect of strategy training on the academic listening of intermediate-level secondary school ESL learners. The experimental group of learners received instruction in two strategies: metacognitive and cognitive.[14] Another group received instruction in the cognitive strategy only and the control group received no strategy instruction. Post-test results showed that the first group and the second group outperformed the control group although the differences approached but did not reach significance.

Further evidence for the role of strategy training in listening comprehension comes from a longitudinal study by Thompson and Rubin (1996). Thompson and Rubin trained an experimental group of 24 English speaking university students learning Russian to use metacognitive (e.g. planning, defining goals, evaluating) and cognitive strategies (e.g. predicting content based on visual clues, background knowledge, genre of the segment) while listening to video-recorded text over a period of two years. Experimental subjects' performance measured by comprehension tests containing open-ended, guided recall and multiple-choice questions was compared to a control group who did not receive any strategy training in listening to video-recorded text, and they showed significant gains in their listening comprehension ability.

---

[14] See definitions provided earlier.

Strategy training studies discussed above showed that L2 learners can benefit from such instruction. However, these studies might be the only which showed an unambiguous effect of strategy training. As Field (1998) notes, reviews of strategy training by Rubin (1994) and Chamot (1995) could only identify two strategy training studies with clear positive effect of a strategy-based approach. In addition, a closer look at the assessment of comprehension measures used in most studies examining the effect of top-down cues or strategy training points to measures being mainly either recall or summary tasks. While these tasks may show that learners have got the meaning and the message of the aural input, they by no means show that learners have managed to learn the L2. Therefore, if listening teachers are helping learners get the message with few perceptual cues, learners are unlikely to develop bottom-up skills which are necessary to learn the language as a grammatical system.

Field (1998) also cautions against over-reliance on a strategy-based approach. His main claim is that strategies or top-down processing skills which can transfer from the L1 should be used only to compensate for the non-automaticity of EFL bottom-up skills. In other words, strategies should not be used to substitute for bottom-up skills (ibid). Rather, as Field maintains, the focus should be on developing bottom-up skills, which native speakers have automated and non-native speakers need to automate. The more automatic EFL learners' bottom-up skills are, the less the need for compensatory cues (ibid). Therefore, the predominance of strategy-based or top-down training in EFL/FLT will reduce learners' chances of developing their bottom-up skills as they will become more dependent on these compensatory cues instead of attending to the actual aural input. Rost (2001: 110) also notes that "listening tasks and instruction should aim to help learners' automise 'lower-level' processing of language so that they can devote more attention to 'high-level' goals". Therefore, it is reasonable to say that a strategy-based or a top-down processing training seems to be avoiding the real source of EFL listeners' problem and puts the cart before the horse (Norris 1995).

Unfortunately, the fact that L2 listening comprehension instruction has been dominated by top-down processing training in FLT is widely recognised (see e.g. Cauldwell 1996; Field 1998, 2003; Norris 1995; Rost 2001; Wilson 2003). Despite the positions taken above, there seem to be some reasons for this dominance. To begin with this might have been motivated by some researchers' claims that native speakers' (NSs) do not have problems when listening to running speech because they do not only depend on the perceived signal (Henrichsen 1984: 106), but rather make better use of the context.[15] In Dalton and Seidlhofer's (1994: 26) words, they are more capable of top-down processing. Findings from empirical work (e.g. Hansen and Jensen 1994; Hildyard and Olson 1982; Shohamy and Inbar 1991) have more or less concluded that whereas low-level listeners rely more heavily on bottom-up cues, high-level listeners are better at using top-down cues. This has presumably led practitioners to conclude that they should focus on training top-down skills.

There is a contrary line of research which suggests that low-level and high-level L2 learners are both weaker in bottom-up than top-down processing (Filología Inglesa (a non-dated electronic article).[16] Other evidence suggests that low-level learners sometimes depend more on top-down processing (Field 2004). L2 listeners may, in

---

[15] Although Henrichsen here refers to syntactic context, I am not actually convinced that native speakers' superior listening ability is solely due to better use of any top-down cues.

[16] http://www.publicacions.ub.es/revistes/bells12/PDF/art14.pdf, 21 May 2006.

fact, place more confidence in top-down information and could therefore be misled (Field, 1997; Tsui and Fullilove 1998).

In addition, bottom-up processing seems to be a better discriminator of performance than top-down processing at both levels of proficiency (Tsui and Fullilove 1998). In a large-scale study which extended over seven years in which about 20,000 Chinese speaking L2 learners of English were tested each year, Tsui and Fullilove (1998) investigated which type of processing discriminates low and high-level learners. They used two types of texts: the matching and the non-matching schema type (examples below). In the first type, the schema activated by the initial linguistic text is congruent with the subsequent text. Therefore, subjects were assumed to be able to use top-down processing to arrive at correct answers to comprehension questions presented prior to the text. In the second type (the non-matching schema) the initial schema is refuted by subsequent text and therefore subjects could not rely on guessing based on initial schema but needed to be able to process the subsequent input bottom-up automatically to revise the initial non-matching schema and arrive at the correct answer.

In the non-matching schema text type, for example, the wrong schema is activated by the question (e.g. what normally saves people from fire) provided to subjects prior to the task. After that subjects listen to the text, which also starts with the sentence "Firemen had to work fast to put out a fire which broke out at the Kowloon Tong temporary housing estate". However, subsequent text contradicts this schema as it shows that the fire was put out because of the change in the wind direction. Therefore, the answer to the question asked "what saved the estate from burning down?" was assumed to differentiate those relying on top-down cues only and those paying attention to bottom-up cues and therefore end up revising their initial non-matching schema.

Tsui and Fullilove found that less skilled learners were as good as more skilled learners in getting the answers for the matching schema text right. However, more skilled subjects answered the questions on the non-matching schema correctly more often than less skilled subjects. Tsui and Fullilove rightly interpreted their findings as evidence that unlike more skilled learners, less skilled learners are weaker in bottom-up skills and therefore they tend to rely more on top-down cues and may be misled.

Despite evidence from studies like Tsui and Fullilive's, it has often been assumed, as in the top-down and strategy-based research discussed above, that low-level learners' listening problems are the result of their reliance on bottom-up processing. In other words, ineffective listeners are those who concentrate on bottom-up skills by listening for each word (e.g. O'Malley et al. 1989). If reliance on bottom-up processing is a problem, then it is so because, unlike native speakers, L2 learners are unable to process bottom-up automatically and not because native speakers use top-down cues better. If we do straightforwardly accept the claim that low-level or "ineffective" listeners' problems are because they concentrate on bottom-up processing more but tend to be slow, then one way to help them is to improve their use of bottom-up processing instead of completely shifting their focus to top-down processing. But what has happened in L2 pedagogy has been the contrary. In what Norris (1995: 47) calls "putting the cart before the horse", teachers have tried to simulate native speaker listening situations, by providing their students with vocabulary lists before tasks and all contextual information that a native speaker would usually know in advance in such situations. As Cauldwell 1996: 522) puts it

> The importance of perception of words has been undervalued. The assumption is that perception will be aided by correct interpretation of the context ("these are the words we would expect in this context") or that perception is not necessary, because contextual clues point so clearly to the speaker's meaning.

Consequently, students are told to try to guess the meaning without having to listen to every word. In addition, as was discussed, top-down cues, including visual ones, were investigated and their importance emphasised (e.g. Ginther 2002; Herron, Cole, York, and Linden 1998; Jones and Plass 2002; O'Malley et al. 1989). The problem with such pedagogical practices is that they imply that teaching bottom-processing skills is impossible. Typical of this view is Ridgway (2000), who supports a meaning focused-approach and suggests that learners should be given abundant practice in listening and left to their own devices to acquire the bottom-up skills. Although some bottom-up skills can perhaps be implicitly acquired, in my view, such a proposal underestimates the role that the EFL classroom can play in at least accelerating the rate of acquisition of these cues.

While top-down processing clearly plays an important role in both L1 and L2 listening, as discussed above, it is difficult to ignore the problems which come with Henrichsen's and Dalton and Seidlhofer's proposals, that native speakers' superior listening ability is due either totally or even mostly to their better use of top-down processing skills. Below, I will argue that this proposal has two problems. The first point discussed in the next paragraph presents the problem which relates to their proposal's implausible conclusion. The next points discuss why the pedagogical implications derived from this conclusion, if correct, can not be applied to L2 listening.

First, as will be discussed in detail in Chapter 3, psycholinguistic research has shown that native speakers use bottom-up processing in performing linguistic tasks such as recognising phonemes, words and identifying the rhythmic structure of a language. In addition, it has been shown (e.g. McQueen 1998) that native speakers greatly depend on bottom-up cues from the speech signal in arriving at word boundaries. Lexical segmentation is known to cause great problems for non-native listeners (Altenberg 2005a; Field 2003) and is the issue addressed in the present study. Therefore, claiming that native speakers are better listeners only because they use top-down processing better than non-native speakers is inaccurate if we realise that segmentation in the L1 is aided by bottom-up cues in the signal.

Second, training L2 learners to use top-down processing skills in L2 listening can not always help as there are some situations in which native speakers, let alone L2 learners, can not rely on background or contextual information because there is none or very little. Take an academic setting, for instance, in which new specialised topics are presented to students. What are listeners' chances to use background or contextual information? In such situations listeners have to rely mostly on what they actually hear as they can not make much use of background information if they are not specialists in the topic they are listening to or if they are new to the speciality.

Third, it has been claimed (Field 2003: 325) that because of the listeners' need for some perceptual information before any contextual or background information can be drawn on, misperceiving the initial aural signal can lead to the use of the resulting incorrect contextual information. Field gives the example of an L2 listener mishearing *I won't go to London* as *I want to go to London,* which produces a set of expectations for what will come next. These expectations, as Field claims, are difficult to revise for L2 listeners even when these expectations are contradicted by new evidence from the text. One reason for this might be because L2 listeners, especially lower-level ones, are weak in bottom-up skills. Therefore, they tend to trust more and stick to whatever information top-down processing may reveal to them, a suggestion that is empirically supported (Field 1997; Tsui and Fullilove 1998).

Lastly and importantly, overstating the role of top-down processing and paying scant attention to improving L2 learners' bottom-up skills ignores one of the most important aims of L2 listening: learning the L2. Let us assume for the sake of argument that by concentrating on top-down skills we arrive at an extreme situation where by relying on top-down cues L2 listeners can understand 100% of the meaning of a text by doing without a large percent of the actual aural input (e.g. using, say, only 10-20% of it). What are these listeners' chances for developing their IL grammar through internalising new L2 syntactic and phonological rules or even vocabulary? An example from an ESL leaner who had at the time lived in the UK for about one year illustrates my point. This learner reported to me that although in the first three months of his arrival he encountered the word *lid* frequently he failed to figure out exactly how it was pronounced. That is because every time he bought a cup of tea or coffee from the university café, he knew that the server would ask if he would like a lid and also pointed to a pile of lids on the counter while asking. He reported that he did not feel the need to listen exactly to how the word was pronounced in order to know what the server was talking about and he could not himself produce the word. This shows that while top-down cues are communicatively helpful they can be counterproductive in the long run from an L2A perspective, particularly if over-relied on. As Rost (2001: 110) emphasizes "comprehension is one of the goals of listening, not the end goal". Therefore, unless the aim of L2/FL listening is purely for simple communication, teaching top-down processing should not be solely resorted to.

In addition, assumptions about the role of noticing in L2A give further support to the role of bottom-up skills where the idea is that the learner must pay attention to particular linguistic features in the acoustic input if these features are to be acquired (Gass 1990, 1991; Hulstijn 2001; Schmidt 1983, 1990, 1994, 1995). It is difficult to argue against the claim that by concentrating on developing L2 learners' top-down skills at the expense of their bottom-up ones we are reducing their chances of using L2 listening as a means of L2 learning. As a result of top-down focus, very little input would become intake under Schmidt's Noticing Hypothesis. Over-reliance on top-down cues in listening may explain why some L2 learners fossilize in L2A of syntax, morphology, phonology or even vocabulary as they may have accustomed themselves to getting the message with the least actual aural input they can manage with. The task of ESL/EFL teachers should be to help students rely less on top-down cues by enhancing their ability in automatic and accurate decoding of input (Tsui and Fullilove 1998).

Having said that, the role of top-down processing should not be underestimated, as successful L2 listening relies on both. However, a more direct (i.e. putting the horse before the cart) approach than top-down processing skill teaching should be considered in the teaching of L2 listening. That is because by concentrating on top-down training, as has been argued, learners are neither assisted in acquiring the language nor are their listening problems addressed. In my view, top-down processing skill teaching is like painkillers that should not be continuously used unless the source of the pain is an incurable condition. Unless substantial empirical evidence suggests that bottom-up skills are not teachable, EFL teachers should teach such skills as these skills are mostly the basis of their students' problems.

2.3.3 Teaching bottom-up

If we teach students bottom-up skills then our aim should be to try to help them process continuous speech automatically. In order to do so, understanding the skills involved at the bottom-up level is expedient. As was discussed in Section 2.2.3, Richard (1983: 228) listed some of the skills involved at the bottom-up level. These include the ability to recognise distinctive sounds, stress patterns of words, words in stressed and unstressed positions, reduced forms of words (e.g. contractions) and recognising word boundaries.

As mentioned in Chapter 1, one of the characteristics of connected speech is its speed. Griffiths (1990; 1992) has shown that the slower the speech rate the better the comprehension by L2 learners. Griffiths (1990) had 15 Japanese ESL learners listen to three passages at three different speech rates: moderately fast, average, and slow. Subjects' answers to 15 true-false questions after each passage revealed that lower scores were obtained at the fastest speech rate suggesting that fast speech rate can hinder comprehension. Building on such research Zhao (1997) investigated the effect on listening comprehension of allowing learners to control the speech rate of the text. Fifteen ESL learners studying in the USA with a level ranging from intermediate to advanced participated in his study. Using a special computer programme Zhao tested his subjects in four different conditions: (1) neither repetition of the passage nor varying the speech rate allowed; (2) no repetition but speech rate can be varied to arrive at an ideal speech rate after which no control allowed; (3) both repetition and varying the speech rate allowed throughout the listening task (albeit on a per-sentence basis); and (4) repetition allowed but speech rate can not be varied. Zhao found that a large percent of his subjects slowed down the speech in the conditions where they were allowed to do so. Subjects' comprehension (measured by comprehension questions) was better in conditions (2 and 3) where they had control over the speech rate. He also found that comprehension was slightly but not significantly better in condition 3, where they were also allowed to repeat the passage, than in condition 2 where they were not.

One pedagogical problem with such a technique, as Zhao concedes, is that if allowed to control the speech rate, L2 learners may tend to stick to a comfortable slow rate and resist the need to proceed to a faster one. Imposing a time limit after which learners have to move to a new rate is proposed by Zhao to overcome this problem. How such time limits can be imposed remains a practical problem because if the option is there, L2 learners are likely to resort to slower speech rates whenever they are unable to cope with fast connected speech.

In addition, the validity of such an approach to teaching EFL listening should be treated with caution. A technique such as Zhao's might be good in allowing learners to recognize words that would otherwise go unrecognized (Vandergrift 2004). However, it is likely that it might consume much of the time which would otherwise have been allotted to practicing listening to normal-rate speech where learners can encounter connected speech phenomena. Another point concerns the technology used to slow down the speech. It is just a matter of time expansion or compression. In other words, if a fast passage of continuous speech with assimilations, reductions etc. is slowed down then these phenomena will still be pronounced. So in a way it is fast speech heard more slowly. This type of speech does not resemble the natural slow speech found in registers such as foreigner talk in which these connected speech phenomena may not be used (see Tarone 1980: 423). A good way of introducing students to reduced forms would be to aurally provide them with a gradual transition from the citation form of words to the phonological processes that these words undergo in connected speech.

Teaching students reduced forms can help them comprehend speech (Brown and Hilferty 1987) as it seems that the presence of reduced forms has an effect on L2 listening comprehension. Henrichsen (1984) investigated how far the presence of sandhi-variation[17] can influence the input-to-intake process. Henrichsen (1984: 104) claims that for input to become intake a process is involved that is governed by several factors. One of these factors is perceptual saliency, which as she maintains, can be reduced by the presence of sandhi-variation, consequently influencing learners' comprehension of the input and the amount of intake they can get. "Assuming that comprehension is dependent not only on the signal clarity but also on cognitive factors" (ibid: 103), she hypothesised that the higher the level of proficiency, the less presence of sandhi-variations can reduces comprehension and vice versa. According

---

[17] another term used to refer to reduced forms (henceforth RFs) and defined as "the phonological modification of grammatical forms which have been juxtaposed" (Crystal 1980: 311 in Henrichsen 1984: 105).

to Henrichsen, that is because listeners with a greater knowledge of the language system are less dependent on the signal itself.

The subjects in Henrichsen's study were 65 students at a US university at three distinct English proficiency levels as follows: one (low ESL 1) with 10 students; another (low ESL 2) with eight students; one (high ESL 1) with 17 students; another (high ESL 2) with 15 students and finally a group of 15 native speakers. Subjects were randomly assigned to one of two different treatment conditions (presence or absence of sandhi-variation). However, because a repeated measure design was employed, each subject later experienced the other treatment condition. Henrichsen used a modification of the Integrative Grammar Test (henceforth IGT) developed by Bowen (1976) to measure subjects' comprehension in the two treatment conditions. The IGT consists of 50 taped sentences which contain some form of sandhi-variation as in 1 and 2 below.

1- Who'd he been to see?
2- Who'd he like to see? (Bowen 1976: 31 in Henrichsen 1984: 111).

Fifteen of the IGT sentences were selected which included second words (prepositions, pronouns, modals etc.) and a variety of types of sandhi-variation (contraction, reduction or assimilation). Two forms of this test, which Henrichsen calls a sandhi-variation exercise (henceforth SVE), were then created: form A and form B, which contained 15 mixed (presence and absence) sentences each. Taking both forms (A and B) with half an hour between them, subjects in the two groups were instructed to listen to the entire sentence then write it down using only the full form.

ESL subjects' results showed that high level learners scored significantly better in the SVE than low-level ESL learners. Moreover, both ESL groups scored significantly better in the absence than in the presence condition. In addition, the mean scores of low-level and high-level ESL learners were not significantly different when sandhi-variation was present but they were when it was absent. For native speakers, there was no significant difference between their scores in the two treatment conditions. However, in the presence condition, NSs scored significantly better than high ESL learners, but in the absence condition there was no significant difference between the two groups.

Henrichsen attributed these results to the adequate knowledge that NSs can rely on when the saliency of input is reduced. On the other hand, because of their limited knowledge of English, ESL learners have to rely more on the input signal (ibid). However, contrary to what she hypothesised, the difference between the absence/presence condition was not the greatest in the results of the low-level learners, rather it was the greatest in the results of the high-level learners. She claims that this finding is a result of the sentence length and complexity even when sandhi-variations were absent, which could be why the scores of the low-level learners were low even in the absence condition.

Since the scores of the low-level learners could have been affected by sentence complexity, Ito (2001) tried to re-examine the effect of RFs on listening comprehension and consequently on input-intake process by modifying two aspects of Henrichsen's study. The first aspect was sentence complexity in the test. The other one was categorising RFs into two types according to the derivation of the form. Whereas phonological forms are derived by the application of phonological rules e.g. *she's* (derived from *she has* or *she is*), lexical forms are not derived by phonological rules but tend to be memorised as one lexical item e.g. *will not* becomes *won't* (Ito 2001: 103). Ito hypothesised that ESL learners' comprehension will be affected differently by the type of RFs, with the more salient (lexicalised forms) affecting them less than the less salient ones (phonological forms).

Subjects were 18 ESL learners from two different levels of listening/speaking classes (nine advanced (henceforth NNSs-upper) and nine intermediate (henceforth NNSs-lower)). Following Henrichsen's method of measurement, Ito (2001) used a dictation test consisting of 20 sentences with RFs. However, these 20 sentences had lower syntactic complexity than those used in Henrichsen's study. Moreover, two types of RFs were incorporated into the 20 sentences, ten sentences with lexical forms and ten with phonological forms. Two versions (A and B) were used, both of which contained the same 20 sentences but in a different order.

Results showed that, as in Henrichsen's study, whereas NSs' absence and presence mean scores were the same i.e. 39.89[18], NNSs' scores were better in the absence than in the presence condition. However, despite modifying the complexity of the test sentences, the effect of RFs on learners' listening comprehension did not vary

---

[18] Maximum score is 40.

according to their level of proficiency. In addition, as Ito hypothesised, NNSs in both groups scored higher on lexicalised than on phonological forms.

Although the two studies discussed above have some limitations that will be discussed later, generally speaking, they show that listening comprehension and consequently the input-intake process can be affected by the presence of RFs. However, the results of these two studies were not identical. Whereas Henrichsen (1984) found that the higher the level of proficiency the higher the effect of RFs, in Ito's (2001) study the interaction effect between the presence of RFs and proficiency level was not found to be statistically significant. Ito (2001: 113) attributes this difference to the fact that "sentences in the dictation test might have been too easy, or too difficult, yielding a ceiling or floor effect, respectively". Had Ito considered the results more carefully, he could have been more specific as to which of the two extremes, too easy or too difficult, resulted in this non-significant interaction effect. A closer look at his results shows that in the presence of RFs, the mean score of NNS-upper was 34.78 out of 40 whereas its NNS-lower counterpart was 29.00. On the other hand, in the absence of RFs, their mean scores were 35.89 and 35.22, respectively. These results show that subjects' results were fairly high in both conditions. This may mean that sentences were too easy, yielding a ceiling effect.

In addition to the fact that the syntactic construction used in the sentences appeared in a popular grammar book for ESL beginners (Azar 1996), another reason for this result might be that, unlike Henrichsen, who used different types of RFs in his test, Ito used only one type (i.e. two categories of contractions). This raises the question of how easy contractions are compared to other types of RFs (e.g. assimilation, elision etc.).

Another important matter that should be discussed here is the method of measurement used in the two studies. Instead of using an oral elicited imitation task, subjects in both studies were asked to listen and then write the full form of sentences where RFs appeared. Henrichsen provides two reasons for preferring this method to the oral elicited imitation task. His first reason was to distinguish the phonologically similar features as in *he is so fat* and *he closed the door* which were indistinguishable from *he so fat* and *he close the door.* In addition, he claims that writing the whole sentence helped "strain short-term memory" forcing subject to rely more on their internal linguistic systems and less on memory (p 113).

The two reasons Henrichsen provides above are questionable. Since the data in this test are not spontaneous, only sentences that do not include phonologically similar features could have been used in the test. Moreover, using this type of test to stretch short-term memory may have an opposite effect which may affect the reliability of the test. Subjects, particularly slow writers, may make mistakes not because they did not comprehend or perceive the RF but simply because they could not recall what they heard. That subjects were asked to write not only the RF but also the whole sentence lends support to this prediction especially when we know that sentences in both tests were 7-11 words long (see appendices of both studies).

If the presence of RFs can hinder learners' comprehension then how far can RFs instruction improve their listening comprehension? To my knowledge, the only study that addressed this question was Brown and Hilferty (1987). Their subjects were 32 adult EFL Chinese students, studying at a language centre in China, drawn from the larger population of all intermediate students at the centre. The subjects were randomly assigned to either a treatment group or a control group.

Selected English RFs, collected and categorised by the researchers, were allocated to four weeks of daily lessons (5-10 RFs each). Each lesson of approximately 10 minutes duration, included presentation of that day's RFs and practice, which consisted of having students respond individually or collectively to questions and statements full of RFs. In addition, seven dictations, each consisting of 20-46 RFs were administered. It was only the treatment group that received RF lessons and dictations. The control group received daily 10-minute drills in discriminating minimal pairs. Brown and Hilferty assumed that such drills would have little effect on subjects' listening comprehension ability

Three measures were used in this study, and each measure had two forms: Bowen's (1976) Integrative Grammar Test explained above, divided into two forms of 50 items each; a version of the UCLA English as a Second Language Placement Examination (ESLPE) listening comprehension subtest, recombined and divided into two approximately equivalent subtests of 25 items each; and two RF dictations, each including 45 RFs presented in the daily lessons. The three tests were administered to both the treatment and control groups at the beginning and end of the experiment.

Analysis of the results showed that the mean differences for the two groups on the three pre-tests were not significant. However, post-test results showed that the means on IGT and RF dictation were significantly better for the experimental group, but there was no significant difference between the groups' means on the ESLPE listening subtest. As Brown and Hilferty (1987: 67-68) argued, had subjects shown improvement only on RF dictation but not on IGT, this improvement could have been completely attributed to the treatment group practicing RFs dictation and the materials in question. Another reason why the improvement in the dictation test could not be completely attributed to the practice effect is that the difference between the means of the treatment and control group was very large, almost twice as much (ibid). These results show that 4 weeks of instruction increased subjects' ability to comprehend and perceive RFs. Presumably, this led to reduction of the effect of RFs working as a filter between input and intake. However, the fact that subjects in both groups' post-test results were the same in the multiple choice listening comprehension test shows the unsuitability of using an integrative test in measuring learners' listening comprehension ability, particularly at the bottom-up level.

As discussed in Section 2.3.3, the main criticism of L2 listening and phonology instruction studies such as Brown and Hilferty's lies in their measures. These measures may not be accurate and reliable in testing the effect of teaching on the target of instruction. As shown above, dictation has been a widely used measure, whether in investigating the effect of the presence of sandhi-variation, or testing the effect of L2 phonology instruction on the perception of L2. Its use as an integrative test in testing listening comprehension was a shift from discrete-point testing (e.g. phonemic discrimination tasks). The basic idea of the discrete-point approach, as Buck (2001) explains, is that units of linguistic knowledge can be isolated and tested separately and based on the knowledge of these units, knowledge of the whole language is assumed. On the other hand, proponents of integrative tests (e.g. Oller 1979) argue that discrete-point tests are not representative of language knowledge as they do not test learners' ability to use and relate these units to each other. The main aim of integrative tests is "assessing the processing of language as opposed to assessing knowledge about the elements of the language" (Buck 2001: 67). Dictation as an integrative test has been argued to assess performance at all stages of speech perception (Oakeshott-Taylor 1977) and therefore it was used as a measure in most of the L2 phonology instruction studies.

2.4 Summary

In this chapter an overview of the processes underlying listening comprehension was presented. Studies investigating listening comprehension processes were reviewed. Empirical studies investigating the effect of different approaches on L2 listening has mainly taken two directions. On the one hand, there are those experimental studies which have tried to improve listening abilities by targeting L2 listeners' top-down processing. While proponents of this methodology argue that it can teach students how to listen, it may rather just teach them how to get the message. In addition, this method was argued to be counterproductive for practical reasons and in terms of acquiring the FL. On the other hand, studies which investigated the effectiveness of adopting a bottom-up approach to teaching listening have not provided clear conclusions regarding the effect of bottom-up teaching. Their measures have often been so general that they fail to pinpoint the source of improvement after teaching.

As discussed in Chapter 1, the current research is of the view that a great deal of L2 learners' listening problems are the result of lack of automaticity in the L2 bottom-up listening skills. Therefore, if we are to investigate the effectiveness of a bottom-up approach in L2 listening teaching we need to target a skill in instruction and then directly test the automaticity of this skill using a well-suited measure. The skill that is the target of instruction in the current study (i.e. lexical segmentation) is discussed in the next chapter.

Chapter 3: Lexical Segmentation

3.1 Introduction

Lexical segmentation, "the process by which the listener divides up the continuous speech stream into linguistically and psychologically significant units that can be used to access meaning" (Cairns, Shillcock, Chater and Levy 1997: 112), is a difficult task for L2 learners (Altenberg 2005a; Field 2003; Goh 2000). That is because unlike the situation with words in a written text which are separated by white spaces, connected speech does not afford the listener with similar reliable cues to word boundaries. Speakers do not pause to separate individual word when speaking naturally. In addition, in connected speech, sounds are assimilated, reduced and elided (Gimson 2001). This results in a great deal of variation in the signal. This takes place not only within words but also post-lexically (i.e. across word boundaries) (Brown 1990).

Consider that native speakers normally do not realise how complicated the task of lexical segmentation is. That is because for them words in fluent speech are apparently discrete and easily recognisable. One reason might be that they have acquired L1 cues that help them locate word boundaries in the signal. However, when we listen to an unfamiliar L2 we realise such cues differ across languages. It soon becomes clear that finding word boundaries is very hard. A naïve L2 listener listening to a sentence in English as in (1) spoken naturally will find it difficult to tell where word boundaries are.

(1) I owe you a yacht.

In connected speech, the only reliable cue for a word boundary would be the pause the speaker may make every five or six words, but this sentence is only five words long. Otherwise, the rest of the speech is a continuous flow of speech sounds. The following spectrogram from Potter, Kopp and Kopp (1966) shows how connected fluent speech is.

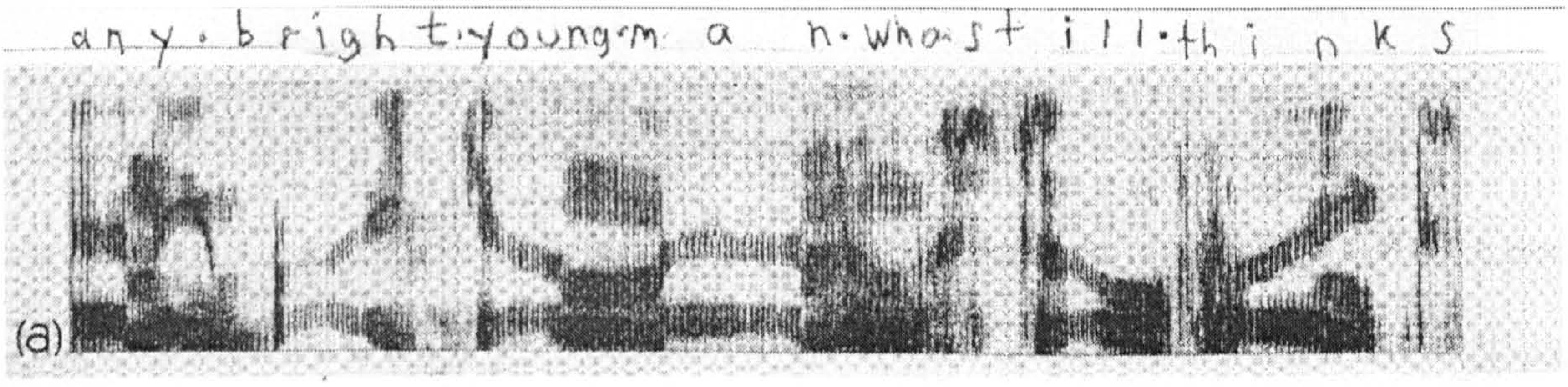

Figure 3.1 A spectrogram of the sentence *any bright young man who still thinks.* From Potter, Kopp and Kopp (1966).

The sound waves shown in the above spectrogram demonstrate that words in the above sentence have no clear boundaries. The beginning of a word is connected to the end of the preceding word and its end is connected to the beginning of the following one. How then does the listener manage to locate word boundaries in order to recognize words? Is this the right sequence for the word recognition process in the first place? In other words, is locating word boundaries what helps us recognize words or does the latter (i.e. recognizing the word) lead to the former?

Competing models of spoken word recognition provide different answers to these questions. Cutler (1996: 87) notes that two distinct approaches to the problem of segmentation have been taken by these models. First, there are models which suggest that segmentation is a byproduct of selecting lexical entries. These are (Serendipitous (i.e. where segmentation is reliant on the likely recognition of words) Segmentation Models (henceforth SSMs). That is, once a word has been recognized the listener can anticipate[19] its offset and consequently knows where the following word will begin. Other models - Explicit Segmentation Models (henceforth ESMs) - claim that, using some cues in the signal, the listener can locate word boundaries even before recognizing the words (Cutler 1996). Both models will be discussed further in this chapter.

This chapter is divided as follows. The next section discusses how researchers have tried to provide an explanation for the segmentation problem using SSMs. At the end of this section the main argument against SSMs will be presented, where we will see how these models fail to explain how the lexicon is acquired in the first place. They also fail to account for the observed segmentation ability of infants, who have yet to acquire a lexicon. ESMs will be presented and supported in section 3.3. Three cues (i.e. allophonic, prosodic and phonotactic) under ESMs will be discussed in sections 3.3.1, 3.3.2 and 3.3.3, respectively. Empirical research on the use of these cues by first language learning infants and adult native speakers will also be presented. Section 3.3.4 sets the context of the current study by highlighting the processes that underlie the learning of phonotactic constraints which are the study's focus. Questions will be raised and answers attempted regarding the language specificity of these cues in section 3.3.5. Section 3.4 then discusses how instruction can help in the L2 acquisition of phonotactic constraints and their use in lexical segmentation. Section 3.5 highlights another form of phonotactics, that is, probabilistic phonotactics "the frequency with which phonological segments and sequences of phonological segments occur in words in a given language" (Vitevitch and Luce 2005: 193). In this section the role that probabilistic phonotactics play in native speakers' processing of

---

[19] I use the word *anticipate* here because some models, as will be discussed below, claim that words are often recognized before their offsets, i.e. before complete acoustic-phonetic information has become available.

spoken words will be discussed. In addition, we will look closely at a proposal of how studying non-native speakers' sensitivity to the probabilistic phonotactics of an L2 may provide answers to debated questions such as how phonotactic knowledge is represented. Moreover, in this section, I highlight some pedagogical implications that could be inferred from the findings of such a study.

3.2 Serendipitous Segmentation Models

As mentioned above, some models of spoken word recognition suggest that segmentation is a byproduct of the speaker's selection of existing lexical entries. However, models under this category differ as to how the lexicon plays that role. On the one hand, we have models which claim that lexical segmentation is achieved through a sequential recognition of lexical items in the speech stream (Cole and Jakimik 1980; Marslen-Wilson and Welsh 1978). On the other, we have models which propose that lexical segmentation is achieved through a process of lexical competition between words beginning at any point in the speech stream. Both models will be discussed in this section.

The first model is one which views lexical segmentation as a sequential process. This view of spoken word recognition and lexical segmentation is best described by Cole and Jakimik (1980: 34) who state that:

> The problem of segmentation is addressed by assuming that speech is processed sequentially, word by word. Each word's recognition locates the onset of the following word, and (along with all preceding words) provides syntactic and semantic constraints on its identity.

One well-known model that proposes this type of segmentation is the Cohort Model (Marslen-Wilson and Welsh 1978). This model relies heavily on the evidence obtained by Marslen-Wilson (1973) that words are often recognized before their offsets i.e. before complete acoustic-phonetic information has become available (Marslen-Wilson 1987). It assumes that on the basis of the sequences of features that make up a word's first 150-200 milliseconds (ms), a cohort of possible words is set up. As more input is processed by the listener, non-matching items are subsequently eliminated from the initial cohort. A word is recognized by the listener at the point when it is the only item left in the cohort (e.g. *slander* will be recognised when /d/ is heard as before that other words are still in the cohort (e.g. *slant*). This is called the uniqueness point. Thus, identifying the offset of the first word, the listener could anticipate the location of the first word boundary. Consequently, another new cohort is set up based on the phoneme that comes next (e.g. when *sland* is heard in connected speech *slander* will be recognised and the phoneme that comes after its offset is assumed to be the beginning of the next word)

As Lively, Pisoni and Goldinger (1994) have noted, this model suffers from some weaknesses. First, the choice of the initial cohort is totally dependent on the listener's correct perception of the initial phoneme. An inappropriate cohort will be selected if for any reason the initial phoneme was not identified correctly (e.g. /s/ was misperceived as /f/ in *slander*). The second problem is that candidate words are assumed to be either in or out of the cohort in a binary manner (ibid). This does not allow for word frequency effects, as will be discussed below.

These observations led to the modification of the original Cohort Model. In the revised Cohort Model (Marslen-Wilson 1987; 1990), the elimination process is no longer all or nothing. Continuous activation functions have been added to the model. Thus items that do not receive further positive input are not eliminated but instead decay in activation. This allows the model to account for the word frequency effects reported in some studies (see for example Savin (1963) and Eimas, Hornstein and Payton (1990)) by assuming that high frequency words receive more activation than low frequency ones (Marslen-Wilson 1987; 1990). This predicts that a low frequency word (e.g. *furl*) will be recognized by the listener more slowly or less accurately than a high frequency one (e.g. *wall*) (see Forster and Chambers 1973 for empirical evidence using a visual lexical decision task).

A critical point in the Cohort Model, however, remains unmodified. This concerns the recognition (uniqueness or isolation) point of words. An ideal recognition point predicts that words can become unique and consequently be recognized before their acoustic offsets. Marcus and Frauenfelder (1985) argue that allowing for variation in the input (unclear signal which might cause an acoustic mismatch) entails that a word is recognized at some stage after it becomes unique. Empirical support for this argument comes from two studies by Grosjean (1985) and Bard, Shillcock and Altmann (1988). Grosjean (1985) used a gating task in which portions of the spoken message are eliminated and in each successive presentation the given message is increased gradually. Subjects try to identify and write down the word after each presentation. Grosjean showed that many short test words presented in sentence context could only be recognized some time after their offsets (e.g. the word *bun* in the sentence "*I saw the bun in the store*" was recognised at the offset of the following word in the sentence).

Bard et al. used spontaneous speech and found that in 20% of the cases words were recognized after their acoustic offsets. Moreover, statistical evidence for this argument is found in an analysis of a lexicon of 20,000 English words by Luce (1986). He found that only about 39% of words are uniquely defined before their offsets.

Goodman and Huttenlocher (1988) tried to investigate the viability of proposing clear isolation points of words (the point at which the non-word deviates from all known words). Using a Lexical Decision Task, in which subjects are presented with real word and non-word stimuli and asked to reject the non-words as fast as possible by pressing a response button, Goodman and Huttenlocher varied when during non-words the isolation point occurred, for example an early isolation point as in *ilvade* and a late isolation point as in *ingade* where the underlined letter represents the isolation point of the non-word. When reaction times were measured from word onset they were faster for early isolation point non-words than late isolation point ones suggesting that non-word identification occurred at their isolation points. However, when reaction times were measured from the isolation point they were longer when the isolation point occurred earlier in the non-word suggesting that information after the isolation point was also needed for identification.

It is embedding (i.e. long words containing shorter ones) that makes any sequential spoken word recognition model difficult to defend. Cutler (1996: 90) points out that subsequent input is necessary to rule out the possibility that a sequence, instead of being a complete word, is only a part of a larger word. She cites the word *fundamentalism* as an example. This word contains words that coincide with its onset like *fun, fund, fundamental* and others that are embedded in the middle like *men* and *mental.*

Using a lexical database, McQueen, Cutler, Briscoe and Norris (1995: 315) found that at least one word is embedded in 83.3% of all polysyllabic words in English and two or more words are embedded in 63.2% of the polysyllabic words. And this is just the situation with words in isolation. In connected speech, as Field (2001: 7) observes, overlap can happen, resulting in a new word appearing between words.

Syntactic cues may not always help in differentiating the embedded word from the word in which it is embedded. McQueen et al. (1995) calculated that 33% of the embedded words match in syntactic class the words that contain them (Cutler 1996: 90).

As shown above, although the sequential account of word recognition exemplified by the Cohort Model may sound like an ideal solution for the segmentation problem by doing without marked word boundaries, statistical evidence shows that this model is not practical, given the high number of embedded words. Short words such as *cap* and *mat* embedded in longer words such as *captive* and *mattress* will always pose a problem for such a model.

Taking this shortcoming into consideration, other SSMs have been proposed (McClelland and Elman 1986; Norris 1994). Like Cohort, these models assume that it is the lexicon that licences segmentation, i.e. that locating word boundaries is a by-product of recognising words. However, they contrast the left-right sequential processing proposed by Cohort and claim that words are recognised as a result of lexical competition between all words that are consistent with the segmentation information of a given input.

One such model is TRACE (McClelland and Elman 1986). TRACE is a fully computational connectionist model.[20] Unlike the original Cohort Model, TRACE does not assume that words are recognized sequentially. In TRACE, there are several processing units or nodes that are arranged into three levels: features (e.g. [+voice], [+nasal], [+sonorant]), phonemes (e.g. /d/, /g/), and lexical entries (words). Nodes at each level of representation are highly interconnected. Feature nodes are connected to phoneme nodes, which in turn are connected to word nodes. Whereas within a single level, the connections between units are inhibitory (i.e. where highly activated units suppress other units at the same level), connections between levels is excitatory and therefore an activated unit at one level can raise or lower the activation level of another unit at another level. In addition, connection between levels is bidirectional. This allows higher level lexical information to influence the activation of phonemes at lower levels.

When a listener hears speech input with the feature [-sonorant, +voice], these features become activated inhibiting all other feature nodes. This activation moves to the next level activating the phonemes that are consistent with these features (e.g. /b/ and /g/). These phonemes then activate the words that are consistent with them. Next, activated words start competing with each other till the highly activated one is recognized. At each stage, activation feeds back to the lower level to reinforce the activation of the consistent nodes at that level.

---

[20] Connectionist models rely on artificial neural networks to simulate and gain information about how natural human processes such as speech processing take place.

TRACE has two advantages. First, its temporal basis solves the problem of embedding because it allows the model to test every segmentation (Frauenfelder and Peeters 1990). Second, its use of higher level lexical information allows it to solve the problem of resyllabification by favouring phoneme segments that create words (Ganong 1980).

Despite these advantages, TRACE is not perfect. Its main weakness stems from its unrealistic treatment of time. As Klatt (1989) notes, it assumes that each phoneme has the same duration but this ignores the temporal variability of speech. In addition, TRACE posits that the entire set of lexical units are duplicated at every third time slice (Field 2001), potentially activating a very large number of words and consequently overwhelming a real listener with a very high number of lexical candidates.

Another model which proposes a lexical competition account of spoken word recognition is Shortlist (Norris 1994). Like TRACE, Shortlist is also a connectionist model. In Shortlist, all words that are consistent with the segmental information in the input are activated regardless of where they begin in the input. The next stage starts when a short list is generated of words that are most consistent with the segmental information. Words in this so-called shortlist start competing with each other for recognition. Finally, words that could provide a complete parse of the input win the competition and inhibit other words that would leave segments unparsed by the hypothetical listener.

McQueen (1998: 22) provides the sequence *enjoyable over-indulgence* as an example to illustrate how Shortlist works. Although candidate words such as *enjoy, enjoyable, joy, over-indulgence, indulgence* and *dull* will be activated, only the words *enjoyable over-indulgence* can win the competition and inhibit other candidates. That is because other candidates such as *enjoy* and *below* or *in* and *dull* will leave segments such as *gence* unparsed.

Shortlist differs from TRACE in that whereas TRACE is interactive (i.e. allows top-down lexical influences on lower units), Shortlist is autonomous (i.e. flow of information is only bottom-up) and therefore higher lexical information can not influence phoneme units. Thus Shortlist supports a modular view of speech processing where a speech module[21] relies on bottom-up information and feedback from the lexicon is unnecessary (Norris, McQueen and Cutler 2000). However, the issue whether speech processing is modular or interactive has been controversial as empirical evidence supporting each argument has emerged. On the one hand, it has been shown that lexical context influences listeners' phoneme identification. Phonemes are monitored faster in words than in non-words (Cutler et al 1987). In addition, when sentences are used, phonemes are monitored faster when words were more easily predicted by context (Foss and Blank 1980). These results were taken as evidence that top-down lexical information influences phoneme identification (but see Norris 1994; McQueen and Cutler 1997 for evidence supportive of autonomous models).[22]

---

[21] See Fodor (1983).

[22] However, as discussed in Ch2, this type of top-down effect is not what most ELT practitioners typically refer to (i.e. background and contextual knowledge) when they call for top-down training.

Lexical competition models of spoken word recognition, whether interactive or autonomous, apparently provide an empirically supported plausible solution to the segmentation problem. Empirical evidence from studies showing word frequency effects (e.g. Forster and Chambers 1973) suggests that processes of lexical competition play a role in spoken word recognition. There is also good evidence that onset-embedded words are more difficult to recognise when they are embedded in longer words (McQueen, Norris and Cutler 1994).

3.2.1 Summary

Three models of spoken word recognition viewing lexical segmentation as a byproduct of lexical identification have been reviewed. The first model (i.e. Cohort) suggested that lexical segmentation occurs as a result of sequential identification of words in continuous speech. This account, as has been shown, is problematic given the statistical evidence of the large number of onset-embedded words. In response to this, lexical competition accounts were then proposed as a solution to this problem. It was shown that these models could solve the problem of embeddings by positing continuous activation of candidate words regardless of where these words begin in the input. Therefore, where a sequential model may incorrectly segment a long word such as *fundamentalism* into the words *fund* and *mental*, this is unlikely to happen within lexical competition models like Shortlist, which will ban such segmentation as it will leave the last part (i.e. *ism*) unparsed.

Although the lexical competition models reviewed have the advantage of identifying onset-embedded words, all Serendipitous Segmentation Models discussed fail to provide a complete account of how lexical segmentation takes place. The problem with any SSM is that it assumes that lexical segmentation of connected speech is achieved using the lexicon. Word boundaries will appear after words are recognized. Although this may sound as advantageous for these models in that they dispense with prior segmentation (Carroll 2004), this assumption fails to explain how a speaker's lexicon is constructed in the first place. As Mehler, Dupoux and Segui (1990) and Cutler (1996) point out, infants lack a lexicon and they need to segment connected speech in order to construct one. Therefore, in not accounting for how the L1A of vocabulary takes place, any proposal of lexical segmentation that relies on the lexicon is implausible.

But is it not possible that infants learn words in isolation in the early stages of vocabulary acquisition and this is what helps them use a lexicon-based segmentation strategy afterwards? This possibility has actually been empirically examined and evidence indicates that this assumption is not correct. Van de Weijer (1998) for instance found that most infant-directed speech contains multiple-word utterances. In addition, there is evidence that most words presented to infants appear in sentential context and that only a very small percentage of words are presented in isolation (Woodward and Aslin 1990). Moreover, Woodward and Aslin found that even when mothers were explicitly instructed to teach their children new words, they only presented 20% of the new words in isolation. There is also the problem of function words (e.g. *the, is, are, will* etc.) (Johnson and Jusczyk 2001). Unlike content words, which might be presented in isolation, the chance of function words appearing in isolation is very low (ibid), yet they are acquired. It therefore seems that infants would have to segment speech using an approach that relies on bottom-up cues available in the signal regardless of what is being segmented (Cairns, Shillcock, Chater and Levy 1997).

3.3 Explicit Segmentation Models

We have seen that SSMs represented as a lexical solution to the segmentation problem fail to explain how infants can segment connected speech without the help of the lexicon. An attempt to provide an explanation that does not rely on the lexicon is made by Explicit Segmentation Models (ESMs). Models under this category use a pre-lexical approach. That is, instead of relying on lexical access in segmentation, they posit acoustic cues that help locate word boundaries before recognizing the words in a sequence. However, it is important to note here that SSMs (particularly the competition ones) are not incompatible with ESMs. Competition models are empirically supported as discussed in section 3.2. The main claim here is that there are many cues at the pre-lexical level that might also signal word boundaries in connected speech.

Different proposals as to the type of the bottom-up cues used have been made. These include allophonic, prosodic and phonotactic solutions. Each of these solutions will be discussed in detail below. Empirical evidence from adult subjects shows that although they have a complete lexicon which might enable them to rely on a lexical-based approach, especially one based on competition, adults are both sensitive to and use the pre-lexical cues in lexical segmentation.[23] However, since I mentioned above that ESMs are meant to explain lexical segmentation ability in infants, I will introduce the seminal work (especially by Jusczyk and colleagues) that investigated the use of these cues by infants in presenting each pre-lexical solution. This sequence of presentation (i.e. infants-older children-adults) is important because it will show us that whereas

---

[23] Hence the importance of teaching bottom-up cues to L2 learners as argued in Chapter 2.

older children and adult can potentially depend on their lexicon in segmentation, empirical research suggests that they continue to utilise bottom-up cues.

First, because infants can not provide explicit verbal or written responses to experimenters, it is important to understand the type of procedures that are used in this regard. These procedures are capable of eliciting the indirect responses from infants to the stimuli. Many of the important research findings about how infants segment words have been obtained by the head-turn preference procedure which was originally developed by Fernald (1985).[24] In the head-turn preference procedure, infants are typically habituated to isolated words (e.g. *nitrate*) and are then presented with passages that do or do not include these words. While listening to the passages, the amount of time that infants turn their head and look at a light they have been trained to look at is recorded. In this type of research, an infant's longer listening to the passage that contains the habituated material is considered to be evidence that the infant is able to segment sound patterns from the stream of speech. I now turn to the first pre-lexical cue used in lexical segmentation.

3.3.1 Allophonic variations

The main point here is that the acoustic phonetic properties of some phonemes differ depending on their phonological context (Church 1987); that is, phonemes have allophonic variations (i.e. allophones) depending on their position in the syllable (e.g. syllable-initial or syllable-final) and proximity to other phonemes. In English for instance, voiceless stops like /t/ and /p/ are subject to aspiration when they are word-

---

[24] For a very detailed explanation of this procedure see Kemler Nelson, Jusczyk, Mandel, Myers, Turk, and Gerken (1995). Also, see Jusczyk (1997: 233-250) for a detailed explanation of the same procedure, the high-amplitude sucking procedure described below, and other relevant procedures used with infants in speech perception studies.

initial and prevocalic and are glottalised when they are word-final (compare *tip* vs. *pit*) (Lehiste 1960 and Church 1987). This means that /t/ produced in words like *top, stop* and *put* is not the same. Thus, if an aspirated /t/ is heard this would indicate that it is word-initial, consequently signalling a word boundary. Similarly, a glottalised /t/ would indicate a word offset.

Other phonetic cues that have been proposed as possible markers of word boundaries include the allophonic variations of /l/ and /r/ (Nakatani and Dukes 1977), the glottal stop, the laryngeal voicing which could mark a vowel as an onset of a new word (Lehiste 1960) as well as the aspiration of voiceless stops (Christie 1974). Christie (1974) and Nakatani and Dukes (1977) will be discussed in detail below. However, we first need to understand what empirical evidence has to say regarding infants' sensitivity to these cues and their use in lexical segmentation.

If infants are to use allophonic variations in lexical segmentation, they should be sensitive to the presence of these variations in the first place. Hohne and Jusczyk (1994) found that infants as young as two months old were sensitive to the allophonic variations of the phonemes /t/ and /r/ in words such as *nitrate* vs. *night rate* which only differed in the allophones of the phonemes /t/ and /r/. In *nitrate* for instance, the [t] is aspirated, released and retroflexed and the [r] is devoiced, indicating that these are word-internal segments, whereas in *night rate* the first [t] is unreleased, unaspirated and not retroflexed[25] and the following [r] is voiced indicating that it is syllable initial (ibid). With younger infants another procedure is used instead of the head-turn preference discussed above. In this procedure, namely the high-amplitude

---

[25] This [t] could also be glottalised.

sucking procedure (Siqueland and DeLucia 1969), infants' sucking rate is measured prior to the presentation of the stimuli using a blind nibble (one without a hole) that is connected to a pressure transducer which is coupled to a polygraph machine. After determining the normal sucking rate for each infant, the experimenter habituates each subject individually to the test stimuli. After that each infant listens to the habituated stimuli and other control stimuli. The difference of the response rates (i.e. significantly greater sucking rate than when listening to the habituated stimuli) is taken as evidence of subjects' sensitivity and ability to discriminate between stimuli items. Using the high-amplitude sucking procedure Hohne and Jusczyk (1994) showed that two-month-old infants could discriminate between the above words (*nitrate* vs. *night rate*). Infants were able to recognise this difference even when prosodic differences were removed through the cross-splicing technique which allows matching the non-critical portions of the stimuli (portions other than /t/ and /r/ in the above example) as regards their acoustic and prosodic differences. [26]

It also seems that this sensitivity is used by infants to lexically segment running speech at an older age. Jusczyk, Hohne, and Bauman (1999a) claimed that infants can use these variations in locating word boundaries. They found that 10.5-months-old (but not 9-month-old) infants were able to perceive the allophonic cues that may discriminate between the items, *nitrate* and *night rate* in fluent speech streams. Using the head-turn preference procedure, they found that infants listened longer to the passage that contained the word that correctly matched the one that they had been familiarized to. That is, when infants were familiarized to *night rates*, they listened significantly longer to the passage containing *night rate* than they did to the passage

---

[26] See Hohne and Jusczyk (1994: 617) for a figure showing the steps of cross-splicing for their stimuli.

containing *nitrate*. This was taken as evidence that by the time they are 10.5-months old, infants are able to actually use allophonic cues in the lexical segmentation of running speech.

Although these results may indicate that infants are sensitive to allophonic variations and that they use them in lexical segmentation, the nature of the procedure used (i.e. sucking and head-turn procedures), although possibly the only usable procedures with infants, may not be so accurate (but see Kemler Nelson et al. (1995) for an argument for the objectivity and reliability of the head-turn preference procedure). Yet findings from infants may be informative when supported by findings from adult subjects with whom more explicit procedures are used. This, however, should not be a precondition for accepting findings from infants that bottom-up cues are used in segmentation. This is because adults, as has been shown above, may use other segmentation strategies (i.e. lexically-based) that can not be used by infants.

The use of the phonetic solution (e.g. allophonic variations) in lexical segmentation by adults may be the least studied solution compared to prosodic and phonotactic solutions. Relatively fewer attempts have been made to investigate if adult native speakers use phonetic cues in lexical segmentation.

One of the studies that found use of a phonetic cue for word boundaries was Christie (1974). Using synthesized speech, Christie compared pairs such as *a star* and *ace tar*. Using a forced choice test, he asked subjects to indicate on their test sheets whether they perceived the stimuli as /a#sta/, where # represents a word boundary, or /as#ta/.

Christie found that subjects used the aspiration of /t/ in the second word to locate the word boundary and identify the word correctly.

Another study which investigated the effect of phonetic cues in lexical segmentation was Nakatani and Dukes (1977). Their main aim was to find out where acoustic cues for word juncture are located. To this end, they constructed hybrids from phrases such as *play taught, no notion, we loan* and *plate ought, known ocean, we'll own.* From these "parent phrases" the transitions from and to the juncture consonant were spliced out and replaced in the original parent phrases in various orders. This was to determine whether the offset of the first word, the onset of the second word or both contributed to the perception of the word boundary. The resulting hybrids were played in random order to 15 secondary school students who were asked to decide from four different options (e.g. no ocean, no notion, known ocean, known notion). Nakatani and Dukes compared subjects' interpretations of these stimuli to determine which part of the stimuli had the strongest cue for word boundaries and found out that the strongest cue for word juncture was always at the onset of the word (see also Christie (1974) discussed above). These cues included aspiration (e.g. /t/ in "play taught"), glottal stop or/and laryngealization of initial vowels (e.g. the first vowel of the word *ocean* in "known ocean") and allophonic variations of /l/ and /r/. The only exception to this rule was with the allophonic variations of /l/ and /r/, which also signalled a word boundary when they were word-final (e.g. we'll own and tour an). In addition, they also concluded that differences in vowel duration (e.g. between *play taught* and *plate ought*, which is long in *play* and short in *plate*) was a weak cue for juncture.

What can we conclude from the results of the last two studies regarding the use of phonetic cues in the segmentation of connected speech? These studies suggest that L1 listeners use allophonic variation when available in lexical segmentation. However, the value of allophonic variation in the segmentation of natural connected speech should be treated with caution. It could be safely argued that even if allophonic variations are used as a segmentation cue they are too weak to be the sole cue. Several factors make it difficult for allophonic variations to stand alone.

The first factor is that there is a limited number of allophonic variations which could help in segmentation and therefore only a small number of word boundaries will be marked by allophonic cues. Another problem with allophonic cues is that they may not mark word boundaries but stressed syllable boundaries (Field 2001). That is because, for instance, voiceless stops in initial position in stressed syllables can be aspirated even if they are not word-initial (e.g. *guit̲ar*) (McQueen 1998: 21).

In addition, Field (2001: 9) argues that allophonic variation can be misleading where resyllabification occurs. So, whereas allophonic cues might distinguish sequences such as *I scream* vs. *ice cream* and *fast steam* vs. *fast team*, cues may lead a listener to choose a false word boundary in a sequence as in (2):

(2) last hour→ [las#tauə].

In the example in (2), because of resyllabification the final /t/ in *last* becomes aspirated. This may lead a listener relying on allophonic cues to think that s/he heard the word *tower.*

3.3.2 Prosodic Cues

Prosodic features have also been proposed as probable cues for word boundaries. Like allophonic variations, infants have been shown to be sensitive to prominence patterns specific to their native languages (Jusczyk, Cutler and Redanz 1993a). In English for instance, the pattern of strong/weak syllables (i.e. syllables with full vowels as opposed to those with reduced vowels, usually a schwa) is predominant and most content words start with strong syllables (Cutler and Carter 1987). This prompted proposing a metrical segmentation strategy in English (Cutler and Norris 1988) which suggests that native speakers of English use strong syllables to locate the beginning of words in the stream of connected speech.

As will be shown below, the metrical segmentation strategy is empirically supported by data both from infants and adult speakers. This indicates that there are certain features which mark stressed syllables in the signal to play that role. Mattys (1997) reviewed several features of stressed syllables. He notes that stressed syllables are:

- Physically salient regarding their pitch, duration and amplitude.
- Phonemically stable in that they are not very vulnerable to phonological modification.
- Perceptually distinctive as shown by different experiments (e.g. Bond 1971; Cutler and Foss 1977) and therefore, they are less vulnerable to misinterpretation than unstressed syllables (Mattys 1997: 319).

Studies have shown that nine-month old American infants appear to be sensitive to the predominant strong/weak (trochaic) stress pattern in English (Jusczyk et al. 1993a). Jusczyk et al. used the head-turn preference paradigm and presented six-month old and nine-month old infants learning English with lists of bisyllabic English words that were equally divided into words that had strong/weak (e.g. gutter) and others that had weak/strong iambic stress pattern (e.g. guitar). Whereas six-month old infants did not show any preference, nine-month old infants listened significantly longer to the strong/weak pattern even when the lists were low-pass filtered.[27] This suggested that infants start learning about the metrical structure of the words in their native language to-be sometime between six and nine months of age.

Infants also seem to rely on prosodic regularities in segmenting words in connected speech. Jusczyk, Houston and Newsome (1999b) conducted a series of experiments to investigate if prosodic patterns help infants in lexical segmentation. In their first experiment, they familiarized 7.5 month old infants with pairs of words with strong/weak pattern (e.g. doctor, candle). When these words where presented to infants in connected speech, they listened significantly longer to passages containing these items than to other passages that contained other items infants were not familiarized to.

To further investigate if infants are responding to the sound patterns of the whole word and not to partial matches involving only the strong syllables of these items, Jusczyk et al. conducted another experiment in which infants were familiarized with just the strong syllables of the words (e.g. dock, can) and then presented with passages containing the whole words (e.g. doctor, candle). In this experiment, infants did not show any tendency to listen significantly longer to the passages containing the

---

[27] Low-pass filtering removes all cues apart from underlying prosody.

monosyllabic vs. bisyllabic words. Jusczyk et al. used this last result as evidence that infants were extracting the whole word and not only the strong syllable.

Thus the evidence provided so far suggests that infants are able to segment words with strong/weak stress pattern from connected speech. If they are doing so by assuming the presence of word boundaries at the onset of strong syllables, then it stands to reason to predict that they will missegment words with the weak/strong stress pattern. This actually what Jusczyk et al. found when they familiarized 7.5 month old infants in the study discussed above with words following the weak/strong stress pattern (e.g. guitar, surprise). Not only did infants fail to recognize these words in connected speech, but also when they were familiarized only with the strong syllable of these words (e.g. tar, prize) they listened significantly longer to the whole weak/strong words *guitar* and *surprise.* In addition, when infants were familiarized with these words in sentential context always followed by *is* and *in* respectively, they listened significantly longer to the isolated pseudowords "taris" and "prizin" suggesting that they were treating them as one strong/weak item.

Taken together the two studies suggest that 7.5 month old infants might be using a metrical segmentation strategy by assuming the presence of word boundaries at the onset of strong syllables (Cutler and Norris 1988; Cutler 1990). Lexical statistics from Cutler and Carter (1987) suggest that this might be to great extent an efficient strategy in English. Using a corpus of 190,000 words of spontaneous British conversation, Cutler and Carter calculated that 74% of strong syllables were sole or initial syllables of lexical words; 11% were initial syllables of grammatical words; and only 15% were not word-initial syllables. Thus the metrical segmentation strategy will lead to correct segmentation in the majority of cases. However, listeners have to correctly segment in all cases and not only in the majority of cases. Relying on the Metrical Segmentation Strategy alone, therefore, can lead to incorrect segmentation when words follow the weak/strong stress pattern.

Although the MSS is efficient to a great extent and the evidence indicates that 7.5 month-old infants solely rely on it when they first start segmenting, for lexical segmentation to always be successful listeners have to incorporate more than one cue to locate word boundaries (Christiansen, Allen and Seidenberg, 1998; Jusczyk 1999; Mattys, Jusczyk, Luce, and Morgan, 1999). In fact, this is actually what 10.5 month-old infants appear to be doing. When Jusczyk et al. familiarized 10.5 month-old infants with the same weak/strong material (e.g. guitar, surprise), they were able to notice the presence of these words presented in connected speech. In addition, infants did not detect the presence of the whole weak/syllable word when they were only familiarized with the strong syllable (e.g. tar, prize) of these words. What is even more important is that, unlike 7.5 month-olds, they were able to detect the weak/strong words presented in isolation when they were familiarized with these words in connected speech always followed by a weak syllable (e.g. "guitar is", "surprise in"). That is, unlike 7.5 month-olds they did not instead detect the strong/weak pseudowords ("taris" and "prizin").

The body of evidence that these studies provided motivated some researchers (e.g. Jusczyk 1997; Jusczyk, Houston and Newsome 1999b; Morgan and Saffran 1995) to suggest that, whereas infants as young as 7.5 months old appear to start lexical segmentation using only prosodic cues, older infants are capable of using other cues (e.g. allophonic, phonotactic) in lexical segmentation.[28]

Evidence from experiments on native English speakers suggests that adults also use an MSS. Using a word-spotting task, Cutler and Norris (1988) found that listeners were slower to spot a word like *mint* embedded in a nonsense sequence consisting of two strong syllables as in [mɪnteɪv], than in a StrongWeak sequence as in [mɪntəv]. Cutler and Norris claimed that because segmentation will occur at the onset of strong syllables, this will leave the listener with the pseudowords [mɪn] and [teɪv] in the first case. As a result, this will delay the recognition process till the input is parsed again and the false boundary is dismissed. On the other hand, because the second syllable in [mɪntəv] is weak, there is no false boundary at the onset of [təv] because a new word will not be assumed at the onset of the syllable [təv] as words in English do not usually start with weak syllables (Cutler and Carter 1987) and therefore [mɪnt] is spotted faster.

Further evidence that native English speaking adults use the MSS comes from a study by Cutler and Butterfield (1992). They found that making slips of the ear involving missegmentation is often (similar to the situation with 7.5 month-old infants in Jusczyk, Houston and Newsome 1999b discussed earlier) a result of either inserting word boundaries before strong syllables where there are not any (e.g. hearing *by loose analogy* as *by Luce and allergy*) or deleting boundaries, where there are ones, before weak syllables (e.g. hearing *how big is it*? As *how bigoted?*).

---

[28] Recall that it has been shown that 10.5 month olds can use allophonic cues (Jusczyk, Hohne, and Bauman 1999a). In section 3.3.3 it will be pointed out that 9 month olds can also use phonotatic cues (Mattys and Jusczyk 2001).

3.3.3 Phonotactic restrictions

Infants and adult native speakers seem to use another cue in segmentation, namely phonotactic constraints which govern the positions in which phonemes and phoneme sequences can appear in particular positions-in-syllable. The role that phonotactics play in lexical segmentation has been widely studied. Generally, two types of phonotactics have been investigated. On the one hand, we have phonotactic constraints: positional and co-occurrence ones. On the other, we have probabilistic phonotactics "the relative frequencies of segments and sequences of segments occurring in syllables and words" (Vitevitch and Luce 1999: 375). Recently, as Gaygen and Luce (2002) have noted, phonotactics has been treated as probabilistic rather than categorical (i.e. legal vs. illegal). Therefore, instead of treating phonotactic constraints and probability separately, they have been merged together to form a kind of a continuum ranging from 'never' to 'very frequently' (Mattys and Jusczyk 2001) or from 'zero probable' to 'high probable'. In this view, listeners can use not only phonotactic constraints which are categorical in lexical segmentation but also the probability and frequency of the co-occurrence of phonemes. Probabilistic phonotactics not only have a role as a segmentation cue, but it has also been shown, using various tasks as will be discussed in section 3.5, that probabilistic phonotactics have various effects on native language speech processing. In this section, I confine myself to discussing phonotactic constraints and their role as a segmentation cue.

Phonotactic restrictions vary cross-linguistically. These constraints govern the kind of phonemes that can appear in a syllable, the positions where phonemes can appear in the syllable (e.g. onset or coda) and which phonemes can be adjacent. The first type of phonotactic constraint could be described as the "vowel constraint" (Brent and

Cartwright 1996: 96). This constraint can be used as a segmentation cue. A listener abiding by this constraint will realize that a sequence like *bigdog* can not be segmented in *b* and *igdog* because this will leave the segmented *b* which does not contain a vowel (ibid: 97).

The second type of phonotactic constraint is the positional constraint. In English for example, /ʒ/ or /ŋ/ can only appear in the coda of a syllable and never in the onset. The opposite is true for /h/. It has been claimed that these phonotactic cues can be used by listeners in lexical segmentation. Hearing /h/ for instance signals a syllable boundary before it. Or does it? Actually this is not always the case. The problem with this type of phonotactic cue is that it relies heavily on the isomorphy of syllable boundaries and word boundaries (Field 2001:10). Therefore, it may not distinguish for example *a hold* from *ahold* (ibid).

The other phonotactic constraint is the co-occurrence constraint. This constraint regulates which phonemes can appear adjacent in onset or coda clusters. In English for instance a labial segment (e.g. /p/, /v/) can not be followed by /w/ in an onset or in a coda when both form part of an onset or coda (Davenport and Hannahs (1998: 147). Similarly, although clusters such as /pr/ and /br/ are completely legal in an onset position (e.g. prime and bread), they are illegal in a coda position. The opposite is true for /nt/ (For a complete inventory of the permissible and impermissible double and triple consonant clusters in onset and coda position in English see Yavaş (2006: 137-

140). Most phonotactic constraints are language-specific.[29] Thus what is illegal in one language may be completely legal in another.

Standard Arabic, for instance, does not allow consonant clusters in syllable onsets. Nowadays, however, SA is the native language of only a small number of Arabs in some parts of the Arab world where SA is spoken at home. Other Arabic dialects that prevail in different Arab countries have phonotactic constraints that do allow clusters.[30]

How can the co-occurrence constraint be used as a segmentation cue? Listeners seem to use the impermissibility of phoneme sequences in finding word boundaries. Clusters such as /pw/ and /vw/ although illegal within syllables in English, can appear between them. The listener assumes that these syllable boundaries coincide with word boundaries. This assumption is always successful as far as monosyllabic words are concerned (e.g. the illegal clusters /pw/ and /vw/) mark out word boundaries as in "*top wing*" and "*move where?*" where *"to pwing or mo vwhere"* would not be heard. However, the experienced listener is able to automatically test the assumption that illegal clusters mark out word boundaries to avoid missegmentation when the illegal cluster only marks out a word-internal syllable boundary rather than a word boundary such as /pw/ in *"upward"*.

Jusczyk (1997) provides another example of how phonotactics of English can be used in segmentation. Knowing the phonotactics of English, a listener will be able to

---

[29] See Section 3.3.4 below for a discussion of the Minimal Sonority Parameter which is claimed to govern what type of consonants may occur in a cluster.

[30] See Chapter 4 for how illegal and legal consonant clusters were identified in the Arabic dialect in question.

correctly segment the sequence *big dog* by assuming the presence of a boundary between /g/ and /d/. That is because the sequence of these two phonemes can not appear in the onset of an English syllable. However, this kind of cue (i.e. where the cluster is illegal only in one position (e.g. /gd/ which is illegal only in onset position but legal in the coda) is problematic as a segmentation cue. That is because this kind of cue is not decisive. In other words, although this cue suggests that the sequence *bigdog* could be segmented as *big dog*, it does not rule out a segmentation like *bigd og* as /gd/ is legal in the coda in English and VC syllables are also allowed in English (see Brent and Cartwright 1996: 97).

Utilising the fact that English and Dutch have similar prosodic characteristics (Crystal and House 1988), Jusczyk, Friederici, Wessels, Svenkcodas and Jusczyk (1993b) tried to find out if American and Dutch infants were sensitive to the phonotactic patterns of their native languages.[31] They presented infants with list of unfamiliar low-frequency words produced by a bilingual speaker. Lists were equally divided into words from the language spoken in the infant's home and words spoken in the other language. The items in the lists from each language had phonotactic patterns that were impermissible in the other language. Jusczyk et al. found that nine-month-old American infants listened significantly longer to words which met the phonotactic constraints of English than to lists of words with Dutch phonotactics. The reverse was true for Dutch infants. Six-month old infants, however, did not show any preference. When the stimuli presented to infants were low-pass filtered, nine month olds did not show any preference. Jusczyk et al. used this result as evidence that infants were

---

[31] This allowed investigators to see if infants were using phonotactic patterns instead of the prosodic ones in identifying the native words.

distinguishing the words based on the phonotactic features and not any remaining prosodic differences.

Not only do nine-month old infants prefer to listen to words that obey the phonotactic restriction of their to-be native language, but they also prefer to listen to non-words with high phonotactic probability (e.g. /tæl/) vs. non-words with low phonotactic probability (e.g. /ʃaud/) (Jusczyk, Luce and Charles-luce 1994). However, as I made clear in the beginning of this section, this sensitivity will be treated separately and therefore will be discussed in more detail in section 3.5.

There is also good evidence that 9-month-old infants can use their sensitivity to phonotactics to segment words from fluent speech.[32] Using the head turn preference procedure, Mattys and Jusczyk (2001) tried to find out if infants actually rely on phonotactic regularities in segmenting connected speech. In their first experiment, 9-month-olds were familiarized with two passages. In one passage, a CVC target word occurred in contexts in which the surrounding words provided good phonotactic word boundary cues. The goodness of the phonotactic cues was estimated from the frequency with which the CC clusters at the onset and offset of a CVC test stimulus (i.e. CCVCC) are found within and between words in child-directed speech, with high between-word probability associated with good cues to word boundaries. In the other passage, the words surrounding the CVC target word lacked good phonotactic word boundary cues. Two CVC stimuli were chosen such that both of their edge consonants could constitute suitable fragments of either within-word or between-word CC clusters with the adjacent words of the passage. A within-word cluster is a CC

---

[32] I will discuss studies showing the use of phonotactic constraints (particularly the co-occurrence constraints) in more detail below as they are closely related to the current study.

sequence that appears frequently within the words of the Bernstein (1982) child-directed corpus and infrequently or never across the words of the same corpus. Conversely, a between-word cluster is a CC sequence that appears frequently across words but infrequently or never within words (Mattys et al. 1999). The two CVC stimuli were the word *gaffe* and the non-word *tove*. In the passage containing the good phonotactic cues for *gaffe,* the CC edges of the word were the between-word clusters /ng/ at onset and /fh/ at offset (e.g. *bean gaffe hold*). In the passage lacking the good phonotactic cues, the CC edges were the within-word clusters /ŋg/ at onset and /ft/ at offset (e.g. *fang gaffe tine*). Similarly, the good phonotactic passage for *tove* involved the between-word clusters /vt/ at onset and /vt/ at offset (e.g. brave tove trusts), whereas the passage lacking the good phonotactic cues had the within-word clusters /ft/ at onset and /vn/ at offset (e.g. gruff tove knows).

After familiarization, infants were tested on repetitions of four isolated items: the CVC target that occurred in the passage with good phonotactic cues, the CVC target that occurred in the passage that lacked good phonotactic cues, and two control CVC items (pod and fooz) that never occurred during familiarization. The trials for the test phase consisted of the four lists of isolated stimuli. Each infant was tested on three blocks of these four lists, for a total of 12 trials. The trials were blocked into groups of four so that each list occurred once per block. The order of the lists within a block was randomized. Mattys and Jusczyk thought that if infants use phonotactic cues to parse connected speech, the target word should be easier to segment from the passage with the good phonotactic cues than from the passage lacking these cues.

Mattys and Jusczyk found that the 9-month-old infants listened significantly longer to the phonotactic-cues present than to the phonotactic-cues absent stimulus passages. This effect was not affected by whether the stimulus was *gaffe* or *tove*. In two other subsequent experiments in the same study, they found that infants showed similar preference for the word that previously occurred in fluent passages even when only one boundary of that word, whether at the onset or the offset, was phonotactically cued.

Results from Mattys and Jusczyk's study provide good evidence of infants' ability to apply their phonotactic knowledge in lexical segmentation. As will be discussed below, similar results were obtained with adult listeners who, have also been shown to be sensitive to phonotactics.

Older children are sensitive to the phonotactic constraints of their L1. Messer (1967) presented 3;7 year-old children with pairs of monosyllabic non-words. One of each pair was a legal non-word (i.e. did not violate English phonotactic constraints) and the other was illegal (either having an illegal initial consonant cluster or both illegal initial and final consonant clusters). Children were asked to say which one of the pair sounded more English-like. Messer found that the possible non-words were judged English-like more often than the impossible non-words. In addition, the impossible non-words were mispronounced more often than the possible ones.

Adult native speakers' sensitivity to the phonotactic constraints of their L1 has also been shown using various tasks. Phonotactic knowledge has been shown to affect speech sound identification by causing misperception of illegal clusters (Brown and

Hildum 1956; Massaro and Cohen 1983; Pitt 1998). In addition, because of phonotactic knowledge, mispronunciations usually follow the phonotactic constraints of the language in question (e.g. if /n/ is mispronounced as [ŋ] it will be in a syllable coda) (Dell, Reed, Adams and Meyer 2000; Warker and Dell 2006). In addition, knowledge of phonotactic constraints can affect subjective ratings of non-words (Coleman and Pierrehumbert 1997). Finally, phonotactic knowledge can cause listeners to hear epenthetic vowels between phonemes when they are presented with consonant clusters that do not conform to the phonotactic constraints of their native language (Dupoux, Kakehi, Hirose, Pallier and Mehler 1999; Dupoux, Pallier, Kakehi and Mehler 2001).

Brown and Hildum (1956) showed that knowledge of phonotactic constraints can cause native speakers to misperceive phonemes. Under conditions of noise, Brown and Hildum presented both naïve and phonetically sophisticated subjects with experimental items and asked them to transcribe them. The sophisticated group who was trained in linguistics was also told to expect illegal phoneme sequences. In creating these experimental items, Brown and Hildum started with ( ) VC context and varied the initial double consonant cluster added to this context. The initial consonant clusters added were either (a) legal in English and resulted in a real English word (e.g. /θr/ in / θrɔːl/ *thrall*) (b) legal in English but did not result in a real English words (e.g. /pr/ in /prɔːl/) or (c) illegal in English and therefore resulted in a non-English word (e.g. /zd/ in /zdrɔːl/).

Brown and Hildum found that only the naïve group identified real words correctly more often than the legal and illegal non-words. However, both groups identified the legal non-words correctly more often than the illegal ones even though the sophisticated group was instructed to expect illegal sequences. This shows that phonotactic constraints can cause a bias towards hearing legal sequences even when subjects were instructed to expect illegal ones.

Similar results were obtained using a different task. Massaro and Cohen (1983) used the phonetic categorization task in which subjects are presented with ambiguous segments in consonant context and asked to report which segment they heard. Massaro and Cohen showed that adult English native speakers' labelling of an ambiguous segment is influenced by their native language phonotactic permissibility of the consonant cluster containing that ambiguous sound. In [t?i] for example the ambiguous segment between [l] and [r] was more often labelled as [r], while the same segment was more often labelled as [l] in [s?i].[33] Clearly, listeners' classifications biased legal sequence (e.g. hearing [tri]) more often than illegal ones (e.g. [tli]) (see also Pitt 1998).

Subjective ratings of non-words are also influenced by phonotactic knowledge. Coleman and Pierrehumbert (1997) analysed native English speakers' acceptability judgement of paired non-words starting with consonant clusters which either respected or violated phonotactic constraints (e.g. /glisləs/ and /mlisləs/). Subjects were asked to judge whether each item could or could not be a possible English word by pressing two response buttons. The number of *no* responses was counted and taken as a score of subjective degree of wordlikeness. Non-words starting with illegal

---

[33] Listeners had to categorize steps along a /r/–/l/ continuum as either /r/ or/l/.

clusters were significantly judged as less English-like more often than those starting with legal ones.

Another way in which native speakers show sensitivity to the phonotactic constraints of their L1 is by hearing epenthetic vowels between phonemes in clusters that do not conform to the phonotactic constraints of their L1. Dupoux et al. (1999) compared the perception of consonant clusters by French speakers vs. Japanese ones. In their first two experiments, they presented their subjects non-words in Japanese and French containing illegal consonant clusters in Japanese but not in French (e.g. *ebzo* where the sequence /bz/ is illegal in Japanese). Subjects were asked to judge if the items contained the vowel /u/ in the middle of the item. The items (n=10) were first naturally uttered, by a Japanese native speaker in experiment 1 and a French speaker in experiment 2, and recorded with the vowel /u/ (e.g. *ebuzo*). After that five more items were created from the original item by modifying the vowel duration. This procedure produced a set of six items. These items ranged from containing a full vowel (i.e. the original item) and no vowel, with vowel duration of 18ms; 36ms; 54ms and 72ms for the remaining four items.

What Dupoux et al. found was that French listeners reported hearing the vowel only 10% of the time in the no-vowel condition and more than 50% of the time only when the vowel length was over 36ms. On the other hand, Japanese listeners predominantly judged that the vowel was present in all conditions even when the vowel was completely removed. The possibility that this effect is merely the result of top-down lexical influence and not because of the phonological context was investigated and ruled out by the same group of researchers 2 years later (see Dupoux, Pallier, Kakehi and Mehler, 2001).[34]

Bilinguals and L2 learners are also affected by both their L1 and L2 phonotactic constraints when listening to their L2. Altenberg and Cairns (1983) asked English monolinguals and English-German bilinguals to rate how English-like visually presented non-words starting with illegal consonant clusters in German only (e.g. smatt) sound. Both groups' ratings were only influenced by phonotactic constraints of English. However, when the same stimuli items were presented in a visual Lexical Decision Task, processing times of the bilinguals were affected by the phonotactic constraints of German as well.

More recently, Altenberg (2005b) examined perception of Spanish-speaking L2 learners of English, with different proficiency levels, of initial consonant clusters. One of the tasks she used was a metalinguistic judgement task in which subjects rate written non-word stimuli starting with legal and illegal consonant clusters in English and Spanish. There were two versions of this task: an English version in which native and non-native subjects rate the non-words as new words of English and a Spanish version in which the non-native subjects rate the non-words as new words of Spanish. The aim of this task was first to find out if Spanish ESL learners have an accurate knowledge of the legality of initial consonant clusters in their L2, and second whether this knowledge was affected by their L1.

---

[34] See also Matthews and Brown (2004).

What Altenberg found was that in the English version of the first task ESL learners were like native speakers in judging non-words starting with illegal clusters in Spanish only (e.g. /snæs/) more English-like than those starting with illegal clusters in English and Spanish (e.g. /sræn/) . As Altenberg pointed out, this result shows that ESL learners from a beginning level have good knowledge of legality of consonant clusters in English. Furthermore, this result shows that non-native speakers' judgements do not seem to be affected by the status of these clusters in the L1. This result as Altenberg noted suggest that the L1 and the L2 "can function independently in a phonological metalinguistic task" (ibid: 73).

Thus it seems from our discussion so far that children, adult native speakers and L2 learners alike are sensitive to the phonotactic constraints of their native language and their L2 and that they show this sensitivity in different ways. We also find that not only are adult listeners sensitive to the phonotactic constraints, but like infants they can use their knowledge of the phonotactic constraints of their native language in segmenting connected speech. In other words, although, unlike infants, adult native speakers may have the ability to use their lexicon to locate word boundaries, they still use bottom-up cues in segmentation. Therefore, the over-reliance solely on top-down training by pedagogical practices in an attempt to simulate native listening situation, as discussed in Chapter 2, is not valid.

Using a word spotting task, Norris, McQueen, Cutler and Butterfield (1997) showed that native listeners take into consideration the phonotactic constraints of what could constitute a phonotactically legal syllable when segmenting speech.[35] In the word spotting task, subjects are presented with aural nonsense sequences where real words are embedded. Subjects have to press a computer key as fast as possible when they spot a word and then say out loud what that word was. Reaction times and error rate (i.e. number of times the target word is missed) are measured in order to see the relation between them and the different contexts where words appear. Norris et al. have demonstrated that English listeners were faster and more accurate in detecting words in nonsense syllables when the remaining segment formed a phonotactically legal syllable than when it did not. Therefore, subjects were faster and more accurate in detecting *apple* in *vuffapple* with *vuff* forming a phonotactically legal syllable than in *fapple* where /f/ is not a legal syllable. Based on these findings, Norris et al. proposed the Possible-Word Constraint (henceforth PWC) as playing an important role in word recognition. However, similar to the vowel constraint but unlike the co-occurrence constraint discussed earlier, the PWC may not be so informative in segmentation. In other words, while the recognition of *apple* in *fapple* might be delayed there is nothing to prevent an infant or an ESL learner to assume that *fapple* as a whole is a word. In this regard, the co-occurrence constraint has the advantage of clearly locating the boundary.

McQueen (1998) is perhaps the first to show clearly that listeners use the sequencing or the co-occurrence constraint in lexical segmentation. He tried to find out if the permissibility of a sound sequence is computed during connected speech recognition and used in segmentation. Using a word spotting task, he asked adult Dutch listeners to spot a CVC word (e.g. pil) in a nonsense bisyllabic CVCCCVC sequence (e.g.

---

[35] This kind of constraint might not be so relevant to the current research, which relies on the co-occurrence constraint. Listeners in this study might be using the phonotactic vowel constraint (see Brent and Cartwright 1996: 97) which states that each word must contain a vowel.

pilvrem). These long sequences had a CC cluster located at the boundary between the target word and the rest of the sequence which was either phonotactically illegal within syllables (e.g. /lv/ in `pilvrem'), thus pointing to a word boundary, or which was very frequent inside of syllables (e.g. /lm/ in pilmrem), and hence disfavouring word segmentation. McQueen thought that if phonotactics are used in segmentation it would be easier to spot a word (e.g. *pil,* (pill)) when it is aligned with a phonotactically determined boundary in a word like *pilvrem* than when it is misaligned with a boundary as in *pilmrem.*

Forty monosyllabic Dutch words were selected to appear in the initial position of bisyllabic nonsense strings (such as *pil* (pill) in [pil.vrem]). Each of these Dutch target words always appeared in a strong syllable, and had four different following contexts. In one context, the following syllable was strong and the phonotactic constraints of the consonant sequence between the two vowels required a syllable boundary aligned with the offset of the target word (StrongStrong, Aligned, as in [pɪl.vrem]). The second context was identical, except that the phonotactic constraints of the consonant sequence demanded a syllable boundary which was misaligned by one segment with the offset of the target word (StrongStrong, Misaligned, as in [pɪlm.rem]). The third and fourth contexts only differed from the first and second in that the full vowels in the second syllables were replaced with the weak vowel schwa (StrongWeak, Aligned,as in [pɪl.vrəm], and StrongWeak, Misaligned,as in [pɪlm.rəm]). A further forty monosyllabic context words were selected to appear as targets in the final position of bisyllabic nonsense strings (such as *rok* (skirt), in [fim.rok]). As with the initial targets, each final target appeared in four contexts, depending on whether the preceding context contained a strong or weak vowel and whether the phonotactic constraints of the sequence of consonants between the two required a syllable boundary aligned with the onset of the target word, or misaligned by one segment with the target onset (e.g., Strong-Strong, Aligned, [fim.rok]; StrongStrong, Misaligned [fi.drok]; WeakStrong, Aligned,[fəm.rok]; and WeakStrong, Misaligned, [fə.drok]).

In analyzing the results, all manual responses (i.e. reactions by pressing the response button) which were accompanied either by no oral response or by a word other than the intended target were treated as errors (5.7% of all responses). Reaction times outside the range of 100 to 1500 ms were also treated as errors (a further 4.6% of the data). Mean error rates (percentage of targets missed) and mean RTs (measured from target offset) are shown in Table 3.1.

Table3.1: Mean percentage missed targets (errors) and mean reaction times for correct detection (RT, in ms), measured from target-word offset in McQueen (1998).

Measure	Target Position	Metrical Structure	Aligned	Misaligned
errors	initial	StronStrong	32%	57%
		StrongWeak	38%	59%
	Final	StronStrong	21%	56%
		WeakStrong	19%	63%
RT	initial	StronStrong	766	828
		StrongWeak	750	809
	Final	StronStrong	535	629
		WeakStrong	499	641

Statistical analysis of these results showed significant strong effects of phonotactic alignment in error rate as well as RTs. Word-spotting was more accurate and faster when words where aligned with phonotactic boundaries than when they were misaligned. In addition, analysis showed that the effects were stronger for final targets where phonotactic boundaries preceded the target's onset. Trying to explain the latter result, McQueen claims that misalignment with a syllable boundary at the word's onset had a strong effect because it could hinder initial access of that word. On the other hand, misalignment at the word's offset can only influence the recognition of an already-accessed word. The overall results were taken as evidence that "the legality of sound sequences is computed during recognition and used to help solve the segmentation problem" (ibid: 36).

The evidence provided so far suggests that infants as well as adult native speakers are capable of using their phonotactic knowledge in lexical segmentation. Further evidence on the role of phonotactics comes from some statistics performed by Lamel and Zue (1984). They found that only 20% of 7000 consonant sequences that potentially occur across word boundaries in English can occur word-medially. In addition, they found that 80% of cross-boundary consonant sequences allow only one division. That is, the phonemes in a sequence such as /mgl/ can only be divided as /m#gl/ but not as /mg#l/ where # represents a syllable boundary (Field 2001: 10).

3.3.4 How are phonotactics learnt?

In the previous section we presented an overview of how adult native speakers respond to non-native sequences that do not conform to their L1 phonotactic restrictions in different ways. L1 phonotactic knowledge seems to force native speakers to either try to assimilate non-native consonant clusters into native ones (e.g. Massaro and Cohen 1983; Pitt 1998) or adopt a repair strategy by inserting epenthetic vowels (Dupoux, Kakehi, Hirose, Pallier and Mehler 1999; Dupoux, Pallier, Kakehi and Mehler, 2001; Matthews and Brown 2004). In addition, phonotactic knowledge seems to play an important role in segmenting connected speech. The question that should be asked here is how this phonotactic knowledge is acquired?

Two answers have been proposed by two different areas of linguistics, namely generative linguistics and psycholinguistics. Whereas both positions posit positive evidence or linguistic experience (the term that is usually used in psycholinguistics) as the main force for learning phonotactics, they differ as to how input is used by learners.[36] On the one hand, generative linguistics suggests that any type of linguistic learning is innately guided by the Principles and Parameters of Universal Grammar (Chomsky 1975). Within this framework, learning language-specific syntactic features is a matter of setting (in L1) and resetting (in L2) parameters to proper values based on the input received. Certain aspects of primary linguistic data play the role of triggers for the setting process.

With regards to phonotactic constraints, particularly consonant clusters, a multi-valued parameter has been proposed (Broselow and Finer 1991). Using the Sonority Index (Selkirk 1982), the parameter proposed, namely the Minimal Sonority Distance Parameter (henceforth MSD) is said to govern what type of consonants may co-occur in a cluster. Co-occurrence of phonemes is constrained by a scale based on the sonority hierarchy (see figure 3.2) on which segments could be rank ordered.

Sonority Hierarchy

Obstruents – Nasals – Liquids – Glides – Vowels

Least sonorous ....most sonorous

Figure 3.2: Sonority Sequencing Generalization (from Broselow and Finer (1991: 37))

---

[36] In generative linguistics positive evidence refers to the naturalistic input in the form of primary linguistic data "which consists of contextualized utterances in the language environment of the acquirer" (Schwartz 1993: 148). This type of input is claimed to be the only relevant data for acquisition. Additionally, it is the most direct means through which learners can form linguistic hypotheses (Gass 2003: 226).

With the most sonorous segment (i.e. vowel) in the centre of the syllable (i.e. the nucleus) values of the parameter determine how close on the hierarchy adjacent segments can be. Since all languages that have double and triple clusters in their onsets also have simple onsets, Broselow and Finer like others (e.g. Eckman 1987) assume that the more consonants are allowed the more marked the option is. This way, MSD allows calculating the markedness value of any given cluster. Using a production task, Broselow and Finer found that Korean and Japanese-speaking ESL learners adopt a value of the MSD parameter that is more marked than their L1 value but less marked than the English value.

This view of how phonotactics are learnt may not be shared by cognitive linguists or/and psycholinguists. Research in this domain argues that learning phonotactics is solely input-driven. It is a kind of statistical learning of the sequential regularities in the input. In this framework, phonotactic constraints are implicitly learnt by responding to sequential frequency of segments, stored in phonological processing systems and generalized to new contexts (Dell, Reed, Adams and Meyer 2000; Chambers, Onishi and Fisher 2003; Warker and Dell 2006). Therefore, phonotactic knowledge is the result of listening experience which adds "perceptual information to the language processing system, with the accumulation of these changes resulting in the abstraction of phonotactic regularities" (Onishi, Chambers and Fisher 2002: 20). Citing studies of learning musical sequences, Onishi et al. further claim that this type of learning (i.e. learning sequential regularities) is not confined to language learning.

Following such proposals some researchers have looked at whether phonotactic constraints can be learnt from brief listening experiences. A study that tried to investigate this question with older infants is Chambers et al. (2003). The authors tried to find out if 16.5 month old infants from monolingual American English speaking homes were able to acquire phonotactic constraints that are not present in English from a brief auditory experience. Chambers et al. familiarized eight infants to lists of CVC syllables which followed artificial phonotactic constraints (e.g. /b/ is always an onset and /p/ is always a coda). In the familiarisation stage, infants listened to a sublist of 25 syllables repeated six times in random order. The average listening time was about four minutes. After that infants were tested using the head-turn preference test using novel syllables that either followed the phonotactic constraints of the experiment or violated them. Chambers et al. found that infants listened longer to novel syllables which violated the phonotactic constraints of the experiment suggesting that infant acquired the phonotactic constraints of the experiments during training.[37]

Similar results were also found with adult English speakers. Onishi et al. (2002) tried to find out if adult English speakers can acquire artificial first-order constraints (i.e. positional constraints like those in the previous study) and second-order constraints (i.e. constraints in which consonant position depended on the adjacent vowel (e.g. /bæp/ or /pɪb/, but not /pæb/ or /bɪp/). Forty subjects were familiarised with study lists containing CVC syllables which followed these experiment-wide restrictions. Using a speeded repetition task, subjects were presented with lists of CVC syllables

---

[37] The preference for illegal items here might not be compatible with the results from infants, especially by Jusczyk and colleagues discussed above, which showed that infants prefer to listen to stimuli which conform to the phonotactic constraints and probability of their native language to be. However, as Chambers et al. (2003: 73) state "The direction of the effect, a preference for the illegal items that least resembled the training set, is consistent with prior experiments with novel materials and familiarization phases of similar duration and complexity" as in Hollich, Jusczyk and Luce (2001); Saffran, Aslin and Newport (1996).

that either followed or violated the experiment-wide constraints. They were asked to repeat them as quickly and accurately as possible. What Onishi et al found was that legal syllables were repeated faster than illegal ones. This result was taken as evidence that subjects acquired the phonotactic constraints of the experiment.

Not only are adult speakers able to gain sensitivity to artificial phonotactic constraints from brief listening experience but it seems that they can acquire sensitivity to the phonotactic constraints of a real language from as short as a seven-minute exposure. Gullberg, Dimroth and Roberts (2007) exposed Dutch listeners to seven or fourteen minutes of naturalistic input in the form of a weather report in Mandarin Chinese. Subjects were then given a lexical decision task. The findings from the lexical decision task showed that whereas this minimal exposure enabled subjects to reject syllables like *gam* as impossible Mandarin words, they were not able to recognise that syllables with consonant clusters (CCCV, VCCC) are not Mandarin.

Gullberg et al.'s results are interesting and relevant to our discussion regarding how phonotactics are learnt. They may suggest a compromise to the two proposals discussed above. On the one hand, these results indicate that learning positional phonotactic constraints entails only statistical learning of the sequential regularities in the input and therefore can be achieved quite fast as was also confirmed by Chambers et al. (2003) and Onishi et al.'s (2002) results. That is because learning this type of phonotactic constraints might be a form of learning phonotactic probability as it resembles learning that a certain segment has zero probability in a certain position. On the other hand, these results may show that learning the co-occurrence phonotactic constraint (e.g. involving the acquisition of consonant clusters) requires setting a certain parameter (e.g. MSD Parameter discussed above) to the proper L2 value and therefore requires a longer exposure.

3.3.5 Language-specific cues

So far, good evidence has been presented regarding three important cues used in lexical segmentation. These cues have been shown to be used by infants learning their L1 and native adults alike. However, these cues are language specific. In other words, languages have different variations of these cues. Let us see first how this applies to allophonic variations. For example, whereas in English [p] and [pʰ] are two allophones of the same phoneme /p/, they are two different phonemes in Thai (Davenport and Hannahs 1998: 106). Also, whereas the phonemes /r/ and /l/ are context-sensitive with allophonic variants in English (Nakatani and Dukes 1977; Church 1987), Japanese speakers do not have a distinction between these two phonemes (Cutler and Otake 1994). Similarly, voiceless stops are not normally aspirated when they are word-initial in Arabic.

Allophonic variations can also be dialect-specific rather than language-specific. As Davenport and Hannahs (1998) report, for speakers in some parts of North England and Scotland the pronunciation of /t/ in *top* and *stop* may not differ as /t/ in *top* is not aspirated.

Similarly, languages have different rhythmic regularities. For instance, the metrical segmentation strategy in English discussed earlier is based on evidence that English is a stress-timed language where there is a contrast between strong and weak syllables (Abercrombie 1964). Whereas a strategy based on such a contrast might also work for Dutch (a stress-timed language (McQueen 1998)), it may not work with either French, which is syllable–based (Cutler, Mehler, Norris and Segui 1986; 1992) nor with Japanese which is mora-based (Cutler and Otake 1994). Thus, it appears that a metrical segmentation strategy critically depends on the rhythmic structure of the language in question. For metrical segmentation to be successful, listeners have to (and they seem to) adapt to their L1-specific rhythmic structure (Cutler 1996).

As noted above, phonotactic constraints are also language specific. Whereas clusters like [ps], [pn] and [pt] can be word initial in German, French and Greek respectively, none of these clusters can be word initial in English (Davenport and Hannahs 1998: 105). Phonotactic constraints can also vary among dialects of the same language. As discussed above, whereas standard Arabic does not allow any consonant clusters in onset position, other Arabic dialects allow clusters in this position.

As we have seen so far, all the cues that have been shown to be used in the lexical segmentation of the Ll are language specific. Recall that in section 3.2 it was argued that infants would have to segment speech using an approach that relies on bottom-up cues available in the signal regardless of what is being segmented. Although they are not perfectly equal, as the L2 learner has acquired the Ll phonology,[38] the same argument can apply to naturalistic L2A.[39] Beginning L2 learners can not rely on their knowledge of L2 vocabulary to segment speech. Similar to Ll infants, they have to segment connected speech in order to acquire L2 vocabulary. However, unlike L1-learning infants, L2 learners may use bottom-up segmentation strategies transferred from their L1. This raises important questions: what do L2 listeners do when listening

---

[38] See Carroll (2004: 234-235) for differences between L1A and L2A in this regard.

[39] I use the word *naturalistic* here to differentiate it from *instructed* L2A, because in the latter words are usually taught in isolation.

to their L2? Do they use segmentation procedures suitable for the L2 or do they transfer their L1-specific procedures?

Let us again start with our first segmentation cue (i.e. allophonic variations). A recent study (Altenberg 2005a) has directly addressed this question. Altenberg investigated if Spanish-speaking ESL learners were able to use L2 acoustic-phonetic cues (e.g. aspiration) to segment English continuous speech. Spanish has no aspirated consonants and therefore if these learners show that they use aspiration as a segmentation cue this can not be the result of transfer. Using a forced choice task as in Christie's (1974) study discussed above, adult English native speakers and advanced and intermediate Spanish ESL learners were aurally presented with naturally produced pairs such as *lou stops* vs. *loose tops* and *keep sparking* vs. *keeps parking* placed within the carrier phrase '*say---------again*'. Subjects were asked to indicate on an answer sheet which of the two phrases they heard by circling the correct choice.

Altenberg found that whereas native speakers' mean percentage correct was 96.7% non-native speakers' mean percentage correct was only 58.5%, which is close to the result expected by chance as subjects had only two choices. Therefore, non-native speakers were significantly less capable than native speakers of using aspiration as a segmentation cue. However, as Altenberg noted, while this result shows that non-native subjects were far from native-like performance, the 58.5% was significantly better than the 50% expected by chance. Therefore, according to Altenberg this result indicates that Spanish ESL learners have learnt to sometimes use aspiration as a segmentation cue given that aspiration is not available in their L1. Thus it seems that ESL learners can implicitly learn to use some allophonic cues in segmentation.

Let us now review what the research has to say regarding our second segmentation cue, i.e. prominence cues. First, several studies conducted with monolingual speakers suggest that adult listeners do not show evidence of using a rhythmic segmentation strategy appropriate for the L2 if this strategy is different from their L1 strategy. It has been shown for example that neither English listeners (Cutler, Mehler, Norris and Segui 1986) nor Japanese listeners (Otake, Hatano, Cutler and Mehler 1993) use a syllable-based segmentation strategy when listening to French. Similarly, neither English listeners (Cutler and Otake 1994) nor French listeners (Otake et al 1993) used a mora-based segmentation strategy when listening to Japanese.

If monolinguals apply their L1 prosodic segmentation strategy regardless of whether it is appropriate or not, it seems therefore that this inappropriate use is a major cause of the difficulty that faces beginning L2 learners. This also raises the question of whether L2 learners can acquire and adopt an L2-specific prosodic segmentation strategy. Cutler, Mehler, Norris and Segui (1992) investigated this question with balanced English-French bilinguals. Subjects were grouped according to the language they preferred to keep in case of a brain injury. A target syllable detection task was used to investigate this question. In this task subjects had to detect as quickly and accurately as possible a target syllable (e.g. *ba-* vs. *bal-; pa* vs. *pal* and *sa* vs. *sal*) in French and English words such as *balcon, balance, palace, palmier, salad* and *salvage*. Cutler et al. found that only when listening to French words the French-dominant group spotted the target *ba* faster in *balance* where it corresponds to the first syllable than in *balcon* where the first syllable is *bal* and not only *ba*. The opposite was true for the target *bal*. In other words, the French-dominant group spotted the target *bal* faster in *balcon* than in *balance.* This was taken as evidence that the French-dominant group adopted a syllable-based (suitable for French) strategy when listening to French only. On the other hand, English-dominant bilinguals failed to show a syllabic segmentation strategy either in French or in English. These findings indicate that prolonged exposure to an additional language might not be sufficient in adopting a second specific metrical segmentation strategy. This encouraged Cutler et al. to conclude that in some aspects of speech processing bilinguals might be "functionally monolingual" (ibid: 409) (but see Sanders, Neville and Woldorff 2002 for evidence that a second specific metrical segmentation strategy can be acquired)

However, evidence from the use of phonotactic cues suggests something different from Cutler et al's conclusion. Weber and Cutler (2006) have shown that proficient German ESL learners use both their native German phonotactic constraints and non-native English ones when lexically segmenting speech in English. Weber and Cutler set out to see whether advanced German ESL learners were affected by their L1 phonotactics when segmenting speech in English.[40] Using a word spotting task, they were able to investigate this question by manipulating the phonemes directly preceding the initial phoneme of a particular word to produce four different conditions. As Weber and Cutler note, this allowed them to compare the detection of the same target word in different conditions, thus controlling potential confounds such as word frequency and Neighbourhood Density (discussed in section 3.5 below). These conditions were: (a) a boundary in both English and German (e.g. *pumlock*); (b) a boundary only in English (e.g. *garʃlock*); (c) a boundary only in German (e.g. *marslock*); and (d) a no boundary condition (e.g. *fuplock*). It should be noted that the

---

[40] I will discuss this study in detail as it is the most relevant and similar study to one of the tasks (Word Spotting) I used, see Chapter 4.

no boundary condition here is different from the one in McQueen's study previously discussed. In that study, not only was there no alignment, but also there was misalignment where there should be a boundary between the first and second phonemes of the following nonsense syllable (e.g. *pil in pilm.rem* /mr/ is illegal in Dutch and therefore /m/ will be attached to *pil* therefore delaying its recognition). It can be argued that in this kind of design (i.e. McQueen's) the source of the effect might not be clear. In other words, we do not know whether it was due to a presence of a boundary between the target word and the nonsense syllable or a presence of a boundary within the nonsense syllable which would keep a phoneme attached to the target word.

From CELEX[41] lexical database (Baayen, Piepenbrock and van Rijn 1993) Weber and Cutler selected 36 l-initial and 32 w-initial monosyllabic words as target words. The preceding nonsense syllables, whose last phonemes determined the required conditions, were constructed using long vowels, diphthongs or vowels plus /r/ (e.g. *fum*; *rin*). This way the syllable was phonotactically legal without the coda and therefore "the internal structure of the context syllable did not itself force a particular segmentation" (ibid: 599). Thus, the two segmentations (e.g. /zaɪ#fli:f/ and /zaɪf#li:f/) are both legal in English.

Weber and Cutler arranged the 68 target words with their appended contexts and fillers into four different lists whereby each target word appeared once in the list in a certain condition. Although Weber and Cutler do not report their reasons, this kind of design has two advantages. First, each subject will only have to spot 68 words instead

---

[41] CELEX is the Dutch centre for lexical information. Its homepage can be accessed on http://www.ru.nl/celex/.

of 272 (68×4). Second and more importantly, this controls the effect that repeated exposure to the same word (if the subject had to spot the word four different times) would have on its detection. Weber and Cutler presented their stimuli items to their subjects (48 German-speaking advanced L2 learners of English and 48 native speakers of American English who were used as a control group) using a laptop running the Nijmegen Experiment Set-Up (NESU) experimental software (Baumann et al. 1993). This software measured reaction times from target offset and recorded them. Subjects were also asked to say the spotted word aloud for accuracy results.

Weber and Cutler compared RTs and miss rates (number of times target words are missed) in the first three conditions to those in the *no boundary* condition to find out which condition produced the greatest effect for both listener groups. What they found out was that words in the *common boundary* conditions (e.g. *pumlock*) were spotted the fastest and most accurately. In addition, both groups spotted words in the *English boundary* condition (e.g. *garʃlock*) faster and more accurately that those in the *No boundary* condition (e.g. *fuplock*). Finally, only the L2 German listeners responded significantly faster and more accurately to words in the *German boundary* condition (e.g. *marslock*). Figure 3.3 shows their results expressed as mean percentage reduction in RT and miss rate for each listener group in each boundary constraint condition compared with the *no boundary* condition.

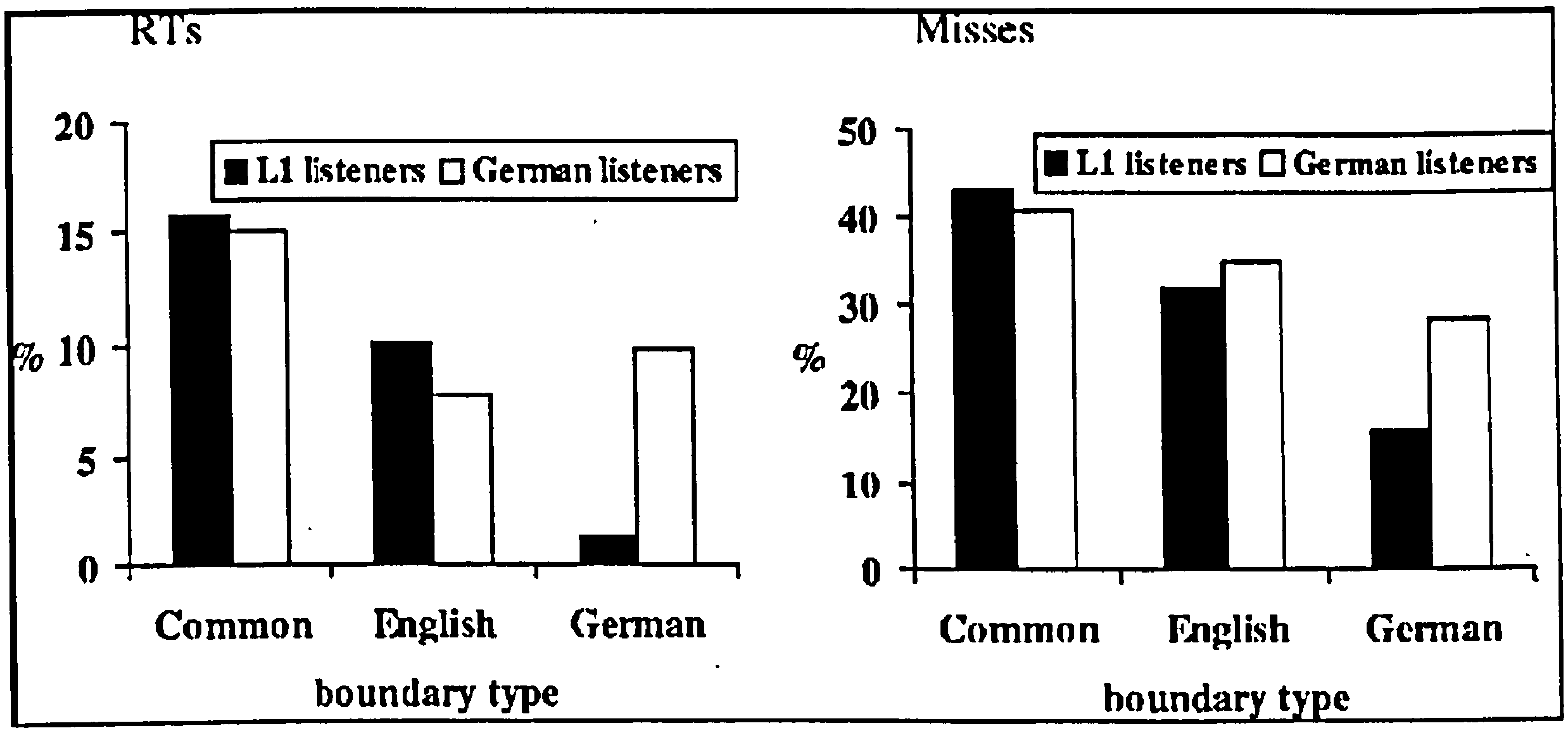

Figure 3.3 Weber and Cutler's (2006) Word Spotting Task results expressed as mean percentage reduction in RT and miss rate, for each listener group in each boundary condition (Common, English, German) compared with the *no boundary* condition (From Weber and Cutler 2006: 603).

These results are interesting. On the one hand, they show that unlike the case with prosodic cues, advanced L2 learners can actually exploit phonotactic constraints specific to their L2 in lexical segmentation. On the other hand, they show that L2 listeners transfer their L1 phonotactic constraints when they are listening to their L2 even if these phonotactics are not informative in the L2 (ibid: 604).

However, these results leave very important questions unanswered. First, the non-native subjects in this study were very proficient, having received an average of 15 years of training in L2 English and therefore their knowledge of English was described as 'excellent'. Thus, their results do not show us how fast L2 phonotactics can be acquired. Do EFL learners at intermediate or lower levels show the same knowledge? Second, this study showed us that L2 learners transfer their L1 phonotactic constraints when listening to their L2 even when these constraints were not appropriate in L2. This latter result raises two questions: first, what role can instruction play in minimising this transfer? And second, what role can instruction play in the acquisition and use of L2 phonotactics in lexical segmentation especially in FL contexts where subjects receive very little naturalistic input?[42]

3.4 Teaching Phonotactics

The previous section discussed the language specificity of three important segmentation cues. Although these cues are language specific, evidence from studies that investigated the use of prosodic and phonotactic cues suggests that adult listeners deploy their L1 cues when listening to an L2. Thus, when cues in the two languages do not coincide boundaries may be falsely detected. The problem is that available evidence from the use of prosodic cues (e.g. Cutler et al 1992 discussed above), suggests that prolonged exposure to an L2 may not enable listeners to use L2-specific segmentation strategies. As a result Cutler (1994) claimed that only one dominant prosody-based language-specific segmentation strategy such as the Metrical Segmentation strategy in English can be used. But Field (2001) has argued against such a claim which suggests that any attempt to train EFL learners, for example, to segment based on the strong syllable is useless as according to Cutler (1994) this EFL-specific segmentation strategy can not be acquired. A conclusion that remains to be investigated.

The present study on the other hand, investigates the teachability of phonotactic cues. Recall that Weber and Cutler (2006) found that, although very advanced EFL learners used English phonotactic constraints, they however transferred their German ones when listening to English. This raises the question of how far explicit teaching of English phonotactic constraints can help less advanced EFL learners improve their

---

[42] This applies to the Saudi context which was discussed in the introduction chapter (Ch1).

lexical segmentation of English, by exploiting English phonotactics and hindering the transfer of L1 ones.

In what way can instruction of phonotactics help in their acquisition? One way is that instruction can play a role in prompting learners' *noticing*, which has been claimed to be necessary in converting input into intake. In his *Noticing Hypothesis* Schmidt (1990, 1994, 1995) maintains that input must be consciously noticed in order to become intake. How can this hypothesis be applied to our current target of instruction? If our Arabic subjects are to acquire the phonotactic constraints of English, they need to notice that certain consonant clusters are absent within words in English. Therefore, it is *Noticing of Absence* which is required.

Highly proficient L2 learners, as Weber and Cutler (2006) have shown, managed to 'notice' the absence of certain consonant clusters in English implicitly (i.e. they were not taught English phonotactics). However, our current EFL subjects, as was discussed in Chapter 1, are not likely to have been exposed to much naturalistic input. Here, the role of instruction is to provide them with a shortcut to acquiring English phonotactics and to aid in lexical segmentation by making the best out of the little input they receive.

In this section the role that phonotactic constraints play in lexical segmentation has been discussed. However, recall that in the beginning of section 3.3.3 I explained that phonotactics has not only been treated as categorical (i.e. legal vs. illegal) but also as probabilistic. The next section discusses phonotactic probability and how studying it can provide answer to both theoretical and pedagogical questions.

3.5 Phonotactic Probability

Phonotactic probability or probabilistic phonotactics which refers to "the frequency with which phonological segments and sequences of phonological segments occur in words in a given language" (Vitevitch and Luce 2005: 193) has been shown to affect lexical segmentation (e.g. Gaygen 1997; Gaygen and Luce 2002). As discussed above, nevertheless, I have made it clear that in the present study, constraints and probability will be treated separately because the main aim of this thesis is to investigate the effect of teaching English phonotactics to learners of EFL on the lexical segmentation of English. Only phonotactic constraints but not probabilistic phonotactics can be taught. Therefore, it was stated that the role of probabilistic phonotactics as a *segmentation cue* will not be investigated in this thesis. However, probabilistic phonotactics not only has a role as a segmentation cue. 9-month-old infants prefer to listen to nonsense words containing high rather than low probability segments and sequences of segments (Jusczyk, Luce and Charles-Luce 1994). Additionally it has been shown using various tasks that phonotactic probability has different effects on adults' speech processing in lexical decision, repetition and word-likeness judgement tasks of an L1 (Bailey and Hahn 2001; Vitevitch, Luce, Charles-Luce and Kemmer 1997; Vitevitch and Luce 1998; 1999; Vitevitch and Luce 2005). This thesis will take previous research a step further by investigating the role that probabilistic phonotactics plays in the *speech processing* of a FL.

One of the studies that asked if phonotactic probability affects the processing of spoken words was Vitevitch et al. (1997). They used two tasks for investigation: a non-word rating task and a speeded auditory repetition one. In the first task, adult native English speakers were asked to use a ten-point scale to rate how English-like the auditorily presented non-words sounded, with number 1 in the scale labelled "Good English Word" and number 10 "Bad English Word". The non-words used were bisyllablic sequences which varied in their phonotactic probability. The measures used to determine phonotactic probability were as follows: (a) positional segment frequency (how often a particular phonetic segment occurs in a position in a word), and (b) biphone frequency (the segment-to-segment co-occurrence probability). In the other task (the speeded auditory repetition task), subjects were asked to repeat as quickly as possible the stimuli items, which were auditorily presented.

Vitevitch et al. found that adult subjects rated non-words with high phonotactic probability as being more English like than those with low phonotactic probability. In addition, in the repetition task, subjects repeated non-words with high phonotactic probability more quickly than those with low phonotactic probability. Vitevitch et al. used these findings as evidence that adult native speakers are sensitive to the phonotactic probability of their native language and that they use this in the processing and recognition of spoken words (see also Bailey and Hahn 2001 for similar findings using the rating task).

However, such findings have proven to be problematic for some competition models of spoken word recognition discussed earlier in this chapter (see also Luce and Pisoni 1998). Recall that those models posited a competition process which takes place among candidates before a word is recognised. In such a framework, Neighbourhood Density "the number of words or neighbours that are phonologically similar to a given word" (Vitevitch and Luce 2005), which is positively correlated with phonotactic probability (Vitevitch, Luce, Pisoni and Auer 1999), is expected to play a negative role in the recognition process.[43] In other words, words in dense neighbourhoods will be recognised more slowly than those in a sparse neighbourhood because they have to compete with a larger number of similar words (Luce and Pisoni 1998). Because of the positive correlation between neighbourhood density and phonotactic probability, stimuli with high phonotactic probability should be processed more slowly than those with low phonotactic probability. How then did Vitevitch et al's (1997) findings show a facilitative effect for phonotactic probability?

This question engaged a number of researchers (e.g. Bailey and Hahn 2001; Vitevitch and Luce 1998; 1999; 2005). One possible explanation was provided by Vitevitch and Luce (1998; 1999). They hypothesised that there are two levels of processing: lexical and sublexical. In this framework, the facilitative effect of high probability phonotactics has a sublexical locus and can only show when non-word stimuli is used, thus taking out the lexicality of the stimuli. In so doing, lexical competition which would have taken place with real word stimuli does not occur and therefore does not slow the processing of the high probability phonotactics stimuli even if these stimuli reside in a dense neighbourhood.

On the other hand, when real words are used, the lexical competitive effects will show and outpower the facilitative effects of phonotactic probability, resulting in a slow processing of the high neighbourhood density stimuli. However, Vitevitch and Luce (1999: 376) stressed that such a proposal does not assume that there are absolutely no

---

[43] Typically, words from dense neighbourhoods have high phonotactic probability and vice versa. That is because words with high phonotactic probability are usually those with many similar words.

lexical competition effects for non-words or that high probability phonotactics can never facilitate processing of real words. Rather their specific claim was that " lexical competition dominates for words, whereas effects of phonotactics are the primary determinant of processing times for non-words" (ibid:376).

Vitevitch and Luce (1998) empirically tested this hypothesis by presenting native English subjects with monosyllabic words and non-words that varied in phonotactic probability and neighbourhood density. Phonotactic probability was determined using the same measures used in Vitevitch et al (1997). Neighbourhood density on the other hand, was computed by comparing a given phonemic transcription (constituting the stimulus) to all other transcriptions in an on-line version of *Webster's Pocket Dictionary*, a 20,000 word on-line lexicon containing computer-readable phonemic transcriptions. A neighbour was defined as any transcription that could be converted to the transcription of the stimulus by a one phoneme substitution, deletion, or addition in any position. Therefore, words such as *tick, pit, piss* and *kick* are all neighbours of the word *pick.*

Two sets of words and non-words were generated: (1) high phonotactic probability/high neighbourhood density stimuli, and (2) low phonotactic probability/ low neighbourhood density stimuli. Using a speeded single-word auditory shadowing task, subjects were asked to repeat stimuli items as quickly as possible. Vitevitch and Luce's findings confirmed their hypothesis. They found that for the non-word stimuli, the same pattern of results in Vitevitch et al (1997) was obtained. That is, subjects repeated high probability/density non-word stimuli more quickly than low probability/low stimuli ones. However, high probability/density real words were repeated more slowly than low probability/density ones.

Vitevitch and Luce (1999; experiment 3) further examined their hypothesis by using a different method. They predicted that if their hypothesis was correct (that there are two levels of processing, lexical and sublexical), using a task that could engage lexical activation for both words and non-words would produce similarity neighborhood effects for both. Such task is the Lexical Decision Task. Since this task requires the correct classification of items into words and non-words, it has been argued that this task entails some lexical processing even for non-word stimuli (Goldinger 1996) in order to arrive at the correct classification. Therefore, the facilitative effect for high phonotactic probability in the non-word stimuli observed in the repetition task (Vitevitch and Luce 1998) will not show. In other words, high probability/density non-words will be responded to more slowly than low probability/density ones. Vitevitch and Luce (1999) presented 240 words and 240 non-words that covaried in phonotactic probability and neighborhood density to twenty native speakers of American English. Their findings matched their prediction. They found that just like real words, non-words with high probability/density were rejected more slowly than those with low probability/density.

Using a variety of tasks, the studies discussed here have shown that probabilistic phonotactics has different effects on adults' speech processing of an L1. It has been shown to affect subjective rating of sequence typicality (Bailey and Hahn 2001; Vitevitch et al. 1997) and speed and accuracy of processing (Vitevitch et al. 1997; Vitevitch and Luce 1998; 1999; 2005). These studies have also shown that phonotactic probability has a significant effect that is beyond and independent of neighbourhood density.

The present study takes these findings a step further by examining L2 English learners' sensitivity to the phonotactic probability of English. Such an attempt is of interest to Vitevitch and Luce's hypothesis regarding the operation of two levels of processing. More specifically, it can provide some insight regarding the source of phonotactic knowledge in a way that native speakers' performance can not. The specific claim I would like to make here concerns the difference between native speakers' and non-native (EFL in this case) learners' lexicons. The difficulty of dissociating the effect of phonotactic probability and neighbourhood density in native speakers lies in large part in the fact that adult native speakers have by definition completely constructed lexicons. The slow processing that takes place when high density stimuli are used is the result of lexical competition of the stimulus item with all similar words in the lexicon. The more dense the neighbourhood the longer it takes to arrive at the correct candidate. But EFL learners do not have as large an English lexicon. Thus, the lexicon, and more specifically neighbourhood density effects, are not expected to play the same role as for native speakers because of the different distribution of words in the two lexicons. In other words, whereas a word in a dense neighbourhood would have to compete with all similar words in a large native lexicon, such competition is expected to be much less in a far smaller non-native lexicon.

The scale of the competition for an EFL learner depends on the number of neighbours acquired by the learner. Common albeit low-density words such as *face, dog, home, book, wife, room* are typically introduced earlier in teaching materials than uncommon albeit high density ones such as *muss, ram* and *soar*. The main point here is that the non-native lexicon is not as informative as a native one as far as its effect on processing is concerned. Therefore, if subjects in the current study (low-intermediate EFL learners) show sensitivity to stimuli that is high probability/ high density in the English native speaker's lexicon, this would provide other evidence that phonotactic probability effects are unique and are not only subsumed by neighbourhood density effects.

Another relevant question that could be answered by the current study design (i.e. using non-native speakers) is the source of information regarding phonotactic probability. When Vitevitch, Luce, Charles-Luce and Kemmer (1996) found that native speakers' subjective rating in a non-word rating task was influenced by phonotactic probability, they wondered about the source of this information. Is it "derived from exemplars of form-based representations of spoken words or is instead abstract knowledge of the probabilistic phonotactic constraints of English"? (ibid: 84). For the same argument regarding the incompatibility of native and non-native lexicons discussed above, similar findings in the current study would suggest that the latter possibility is more plausible.

Such findings would also have some pedagogical implications. Recent evidence (Storkel, Armbrüster and Hogan 2006) suggests that phonotactic probability facilitates learning of novel words (non-words paired with novel objects) by adult native English speakers. In other words, low phonotactic probability items are learnt faster than high probability ones because

> high-probability novel words may be deceptively similar to many other known sound sequences in the language, whereas low-probability novel words will stand apart from other sound sequences as unique. Based on this deceptive wordlikeness, learning may not be triggered

> on first exposure to a high-probability sound pattern but may
> be immediately triggered on first exposure to a low-probability
> sound pattern (ibid: 1188).

Therefore, if EFL learners, similar to native English speakers, are found to be sensitive to the L2 English phonotactic probability, such sensitivity is predicted to have some practical implications. Among these are the speed of learning low-probability new words and the ease and speed of repetition and recall in testing of high probability phonotactic non-words (which can not be distinguished from novel words for EFL learners) as obtained with native speakers (Vitevitch et al. 1997; Vitevitch and Luce 1998; 1999; 2005). Consequently, presentation of new items in vocabulary teaching materials may need to be revised to follow a specific graded order compatible with such findings.

3.6 Summary

Two general views regarding what makes lexical segmentation of connected speech possible have been reviewed. Whereas the first view (Serendipitous Segmentation Models) considered segmentation to be a byproduct of word recognition, the latter (Explicit Segmentation Models), which was supported by evidence from L1A, posited several types of information that help listeners locate word boundaries. Three of these cues (allophonic, prosodic and phonotactic) have been reviewed. Evidence from a series of studies suggested that during the latter half of the first year of life infants develop sensitivities to these cues and start to use them soon afterwards in lexical segmentation. Adults listening to their L1s have also been shown to use these cues. Because none of these cues is sufficient when used alone, it has been argued that for lexical segmentation to be successful these cues must be used in combination.

Although these cues are language specific, evidence (specifically from studies that investigated the use of prosodic cues) indicated that adult listeners deploy them when listening to an FL. Thus, when cues in the two languages do not coincide false boundaries will be detected. The problem is that available evidence from the use of prosodic cues (e.g. Cutler et al 1992 discussed above), suggests that prolonged exposure to an L2 may not in itself enable listeners to use an L2-specific segmentation strategy. Consequently, the current chapter concluded by proposing a role for the explicit teaching of one important segmentation cue, phonotactic constraints. Such teaching has the potential to enhance EFL learners' ability to segment connected speech in English and consequently help their listening comprehension ability.

In this chapter, I also proposed investigating the effects of probabilistic phonotactics on the processing of EFL for important theoretical and pedagogical reasons. However, in the present study, as was clear from the above discussion, constraints and probability are treated separately. That is because the main aim of this thesis is to investigate the effect of teaching English phonotactics to learners of EFL on their lexical segmentation of English. Obviously, phonotactic constraints but not probabilistic phonotactics can be taught. Therefore, the role of probabilistic phonotactics in the processing of EFL, but not as a segmentation cue, is investigated in this thesis.

As stated in Chapter 1, the current thesis aims to investigate four research questions, which for the sake of convenience I repeat here.

1. Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English and Arabic?

2. Are Arabic speaking EFL learners sensitive to the phonotactic probability of English?
3. Do Arabic speaking EFL learners use the phonotactic constraints of English and Arabic in lexical segmentation of running speech in English?
4. Can explicit teaching of English phonotactic constraints help improve Arabic speaking EFL learners' ability in lexical segmentation of English?

Investigating these research questions requires a sophisticated methodological design. Such a design is presented in the next chapter. This chapter also includes more detailed research questions and relevant hypotheses.

Chapter 4: Methodology

4.1 Introduction

As was discussed in Chapter 2, the current study is based on the idea that a great deal of L2 learners' listening problems are the result of lack of automaticity in their L2 bottom-up listening skills. Therefore, it is assumed that a good way of teaching L2 listening is by helping learners automise the use of these cues. Consequently, a good way of testing the effectiveness of a bottom-up approach is by measuring its effect in enhancing the automaticity of learners' performance.

Chapter 2 (Section 2.3.3) reviewed some studies investigating bottom-up listening skills. These studies used integrative tests such as dictation or general listening comprehension tests as measures. However, in the present study (studying the effect of teaching L2 phonology on L2 comprehension), the inappropriateness of using dictation as a measure lies in its assumed advantage in that it assesses so many things). In other words, any improvement in the post-test could not be directly attributed to improvement in the target of instruction as other factors could come into play. Akita (2005), for example, observed in her experimental study, using dictation as a measure of improvement of subjects' perceptual abilities, that improvement in the post-test could not be directly attributed to improvement in the target of instruction. She had three groups: two experimental groups (a segmental group and a prosody group) and a control group. Her aim was to find out which method teaching EFL phonology at the segmental level (e.g. exercises in phonemic discrimination/minimal pairs) or teaching at the suprasegmental prosody level would have a better effect on EFL learners' production and perceptual abilities. Dictation was the measure used to test her subjects' perceptual abilities. After teaching her experimental subjects in both groups for one semester at a Japanese university she found out that all groups, including the control one, had improved with no statistically significant differences found. Such a study shows the difficulty of tracing the source of improvement when dictation is used as a measure.

Moreover, in a dictation test, subjects, particularly slow writers, may make mistakes not because they did not comprehend the dictated sentences but simply because they could not recall what they heard. This is particularly true when dictated sentences are quite long (e.g. 7-11 words in studies by Henrichsen, 1984 and Ito, 2001, discussed in Chapter 2). In addition, the practice effect (i.e. the effect resulting from using the same measure in training as well) may influence the results when dictation is used as an activity during the treatment. The same flaw applies to other integrative tests such as comprehension questions (particulary multiple-choice like those used by Brown and Hilferty (1987) discussed in Chapter 2 (Section 2.33)). In addition to the fact that they fail to pinpoint the source of the improvement, they do not show how much of the text subjects have understood to arrive at correct answers (Field 2003).

Neither discrete-point tests (e.g. auditory discrimination tasks) alone nor some form of integrative tests are proper measures when the effect of teaching certain L2 phonological rules on listening comprehension is investigated. On the one hand, discrete-point tests examine knowledge of certain linguistic units without revealing if this knowledge helps in processing the language. On the other hand, in using dictation as an integrative test, different factors could influence subjects' performance hence resulting in an inability to pinpoint the source of improvement. Therefore, if we teach bottom-up cues we need to test bottom-up cues. In other words, if we are to pinpoint the locus of the effect of a bottom-up approach, we ought to identify a specific bottom-up skill, target it in instruction and then directly measure the effect using an appropriate method.

For the argument presented above, another alternative is proposed here. This proposal takes viewing listening comprehension ability as involving different micro-skills like those proposed by Richard (1983) discussed in Chapter 2 (Section 2.2.3). Such a taxonomy, as Brown (1994) argued, helps teachers develop appropriate teaching methods for each of these skills and also use them as testing criteria. This is actually the methodology that was adopted in the current study. A micro-skill (i.e. ability to determine word boundaries) was targeted in instruction and then directly tested using a particularly suited measure (i.e. Word Spotting Task) as it allows investigating the effect of manipulating boundary conditions on the speed and accuracy of spotting target words.

As was discussed in Chapter 3, the aim of the current research is twofold. The first is to test Arabic-speaking EFL learners' sensitivity to both the probability and the legality of English phonotactics. The second is to find out if they use the legality of English and Arabic phonotactics in the segmentation of English fluent speech and if explicit instruction could help improve their segmentation ability using English phonotactic cues. This entails the use of sophisticated experimental design and special tasks to provide the required data.

Unlike previous instruction studies which used measures that were either so general or metalinguistic in nature, the present study investigated the effect of teaching EFL phonotactic rules on three different levels: explicit, implicit and automatic utilization of the rules. First, I tried to find out if such instruction would result in explicit knowledge of the EFL phonotactic rules which could be applied when subjects are given enough time, hence the use of the Non-word Rating Task. Second, I tried to find out if such instruction would result in implicit native-like knowledge of the EFL phonotactic rules which could result in an implicit automatic sensitivity to these rules, hence the use of the Lexical Decision Task. Finally and most importantly, I tried to find out if phonotactic training would result in implicit, automatic and native-like utilization of these phonotactic constraints in the lexical segmentation of EFL, hence the use of the Word Spotting Task. Using such a design enables us to determine whether non-native speaking subjects' bad online performance, for example, is because they do not know the rules or because they have accurate phonotactic knowledge but are unable to apply such knowledge during on-line processing (Altenberg 2005b).

In what follows, a detailed description of the design and the tasks used is discussed. In addition, an account of how these tasks are to answer our research questions is provided. An overview of the treatment used in the study will also be presented. However, before doing so, an account of the participants will be given.

4.2 Subjects

The nature of the present study required the use of three groups: a native control group to which the non-native groups' results were compared, a non-native experimental group which underwent the treatment (discussed below), and a non-native control group.

4.2.1 Native English speaking control group

12 native speakers of North American English comprised the native control group. Nine of them (seven female, two male) were undergraduate students at Newcastle University as part of an exchange program with American universities. One female participant was a postgraduate student at the school of Education, Communication and Language Science. The remaining two participants were lecturers at Newcastle University and Durham University. All participants excluding the two lecturers were paid £10 each for participating in the study. None had knowledge of Arabic. Before participating in the tasks, each participant was explicitly informed that the tasks were part of a PhD thesis and that their identity would remain anonymous.

Choosing native speakers of North American English as a control group was for a good reason: the input that Arabic-speaking EFL learners participating in this study are exposed to is mainly of North American accented. This is important because in the treatment (discussed in section 4.4.5) subjects were set listening tasks during which they listened to naturally spoken American English input through the media. That is also why the person who recorded the items in the sound files used in the tasks was a (female) native speaker of North American English, thus minimising the possible effect that a different accent would have on processing times and recognition of certain differently pronounced words. What also helped control a potential confound is that calculating the phonotactic probability of our non-words, as will be discussed below, was based on log-frequency-weighted counts of words in North American English.

4.2.2 Non-native groups

The EFL learners who participated in the study were 48 Arabic-speaking male undergraduate students at the Department of English Language and Translation, Al-Qassim University in Saudi Arabia. Worth mentioning is that Al-Qassim province, where the university is located and where students stay at least during term-time, is neither a tourist destination nor industrial and therefore does not attract foreigners including native English speakers. In addition, lecturers at the department are all non-native English speakers. As a result, subjects' chances of mingling with native English speakers during term-time are minimal, thus controlling to some extent the types and amount of input they received.

There were two practical reasons for choosing subjects from this department. Firstly, there was easy access and facilitation of administrative procedures due to the direct connection of the researcher to this department. Secondly and more importantly, recall that spoken Arabic has dialects with different phonotactic rules and phoneme inventories. The researcher's knowledge of the dialect spoken in this area, being his native dialect, was crucial in the choice of the test items used in the tasks, as will be shown in section (4.4.1.1).

The average age of the subjects ranged from 18-20 with an average of 19 years. English study in this department takes four years, which is divided into eight one-semester-long levels with students at level 1 classified as low-beginners and those at level 8 as advanced. Participants in the present study were at Level 3 and therefore classified as low-intermediate. Prior to joining the department, students had already studied English for six years in the intermediate and high school stages. However, due to the methodology (mainly the Grammar-Translation Method) adopted in teaching English at these stages and other educational and social factors, students in Saudi Arabia graduate from high schools with a very low level of English, particularly in terms of their oral communicative ability.

There were two reasons for choosing students at an intermediate level as subjects for the study. First, subjects at a beginning level are likely to be struggling with listening problems resulting from a small EFL vocabulary and insufficient knowledge of the phonological system of English. Therefore, they may have not been ready yet for the new methodology (i.e. phonotactics teaching) adopted, which assumed some knowledge of these things. Intermediate-level subjects are assumed to have the basic required knowledge of EFL. Consequently, their margin for improvement resulting from treatment was assumed to be adequate. These reasons made the intermediate-level subjects a better choice. In addition, subjects were particularly chosen from level 3 because at this level students take a listening module which the researcher was able to take over to give the treatment.

Prior to the pre-tests, two subjects reporting hearing problems were eliminated from the study. Additionally, the pre-test results of three subjects were discarded due to their irregular attendance and those of another three for their incompliance with the listening tasks given during the treatment. This left 40 subjects, who all reported no history of speech or hearing disorder. The 40 subjects who fully participated were equally divided into two groups: a control and an experimental group. Three criteria were taken into consideration in dividing the subjects into these groups: first, the daily average time the subject was exposed to natural English input through the media; second, the subject's grade point average (GPA) scored at the end of level 2; last, the subject's final mark of the listening module scored at level 2. Subjects with similar or identical scores were put into different groups. Table 4.1 shows mean scores of these criteria for the control and experimental groups.

Table 4.1 Non-native subjects' mean scores in the three criteria upon which the division process into control and experimental groups was based.

	Daily exposure to naturalistic input	**GPA Scored at level 2**	**listening module mark scored at level 2**
Control group	7 minutes	3.53 out of 5	78 out of 100
Experimental group	9 minutes	3.61 out of 5	74 out of 100

Subjects participated in the study as part of their Level 3 listening module. However, prior to the pre-tests, subjects were explicitly informed that these tests were part of a PhD study and that participation was optional. In addition, they were also made aware that the study conducted would not affect their official assessment in any way. See appendix A.1 for the consent form subjects signed prior to the experiment, which also included a survey based on which the distribution of subjects into control and experimental groups was made.

4.3 Design

The study followed a pre-test-treatment-post-test design. Tasks that were used in the pre-tests and post-tests are discussed in section 4.4 below. After the pre-tests, subjects in both the experimental and control group were taught by the researcher during an eight-week long term for about one and a half hours a week (see section 4.4.5 below).

Whereas the control group received standard instruction in listening comprehension with no special reference to phonotactics, subjects in the experimental group were taught the same materials as the control groups in addition to explicit teaching and different types of activities with English phonotactics, discussed in section 4.4.5. After the treatment period, subjects were then immediately post-tested with the same tasks. In the following section, a detailed description of the tasks used is presented.

4.4 Task overview

Three tasks were used in the present study, namely Non-word Rating, Lexical Decision and Word Spotting. Whereas the aim of the Word Spotting Task was to test subjects' ability to use English phonotactics in the segmentation of English connected speech, the first two tasks (i.e. Word Rating and Lexical Decision) had one aim. This was to test subjects' sensitivity to the probability and legality of English phonotactics. Using two tasks to answer the same question was important. Recall that in the present study a treatment follows the pre-tests followed by post-tests. As discussed above, using two tasks was necessary to pinpoint the source of improvement, if any, in subjects' sensitivity to the legality of English phonotactics in the post-tests. In other words, it shows us if the improvement is just due to knowing phonotactic rules that can be applied only when enough time is given or whether it is due to an implicit native-like knowledge of these rules which can show in online processing. More specifically, the Word Rating Task, with the time it affords, may allow the subjects (as will be shown in section 4.4.1) to use the metalinguistic knowledge resulting from explicit phonotactic teaching during the treatment in judging illegal non-words. The Lexical Decision Task (discussed in section 4.4.2), on the other hand, denies them a very important condition for metalinguistic knowledge to be used. This condition is time. Therefore, using both tasks was necessary to have some insight into the type of knowledge being used by subjects in judging non-words stimuli in different phonotactic conditions. Is it just metalinguistic knowledge or native-like knowledge that could show as different RTs in responding to different phonotactic conditions in the Lexical Decision Task? In other words, would subjects' performance show signs of automatcity as discussed in Chapter 2? The following sections describe the three tasks used in the study.

4.4.1 Non-Word Rating

Using a scale as in (1), the Non-word Rating Task requires subjects to rate non-words depending on their perceived typicality of the language in question.

1) Very non-typical = 1 2 3 4 5 Real word = Very typical

This task design has been previously used by researchers to study the effects of both phonotactic probability and neighbourhood density (Bailey and Hahn 2001), phonotactic probability and non-word length (Frisch et al 2000) and phonotactic probability and metrical influences (Vitevitch et al 1997) on native speakers' ratings of non-words. It has also been used to investigate the effect of phonotactic legality on subjective ratings of non-word stimuli (Altenberg and Cairns 1983; Altenberg 2005b; Coleman and Pierrehumbert 1997). Unlike other tasks where other factors may come into play, Bailey and Hahn argued that subjects' sensitivity towards sequence typicality of non-words in the word-likeness (i.e. non-word rating) task is the predominant factor for their ratings. Taking into consideration the relatively longer time that this task affords subjects in addition to the unambiguous procedure it follows, this claim sounds plausible.

Another reason for using Non-word Rating along with the Auditory Lexical Decision task in the current study, as discussed in the previous section, was to determine what type of knowledge is being used. That is crucial because we are testing subjects' sensitivity not only to phonotactic probability but also to phonotactic legality, which is rule-governed and teachable.

4.4.1.1 Materials for Non-Word Rating Task

36 monosyllabic non-words were prepared as test items for this task. 12 of these items were monosyllabic non-words starting with illegal onset clusters in English such as /dl/ in /dlaus/. In six of these items, the consonant clusters are illegal both in English and Arabic (henceforth *illegal in 2*) (e.g. /tl/) whereas in the other six items they are illegal only in English (henceforth *illegal in 1*) (e.g. /mr/). In order to control for the effect that the phonotactic probability of the syllable starting at the second consonant of the non-word (e.g /laus/ in /dlaus/) may have on results the phonotactic probabilities of these syllables were always low. See appendix A.3 and A.4 for a complete list of these items along with statistics of phonotactic probability for the syllable starting at the second consonant of the non-word.

As was discussed in Chapter 1, standard Arabic (SA) does not allow two-member consonant clusters in syllable onsets. Yet, SA is the native language of only a small number of Arabs and dialects in various countries may allow clusters. However, in most Arab countries, SA is learnt at schools from an early stage. Additionally, Modern SA, a variety of Arabic that follows most of SA phonological rules, is used in religious contexts, formal lectures and the media. Thus whereas Arabic speakers speak different dialects as their L1, their knowledge of SA and MSA may vary depending on exposure. Thus, our EFL learners' knowledge of SA and MSA was taken into consideration when preparing the test items. Since SA and MSA do not allow any initial consonant clusters, they were ruled out as potential confounds in preparing the test items starting with illegal consonant clusters. On the other hand, subjects' native dialect (i.e. Qassimi Arabic (henceforth QA))[44] was taken into consideration. But since this dialect is understudied, certain procedures were followed to arrive at the illegal consonant clusters in it.

Initially, the researcher collected 25 initial two-member consonant clusters. Some of these clusters were illegal and others were legal in English based on the inventory provided in Yavaş (2006). Being his native language, the researcher could identify 13 of these initial clusters (six were illegal in English and seven were legal) which were used word-initially in QA words. Nevertheless, this did not guarantee that these clusters were perfectly legal in the QA dialect. When pronouncing these words, epenthetic vowels may be used to break up these clusters. As a result, another procedure was used. The researcher suffixed a nucleus and a coda to each onset cluster, producing 13 non-words such as /glauθ/ and /mreʃ/ (see Appendix A.2 for a complete list). These 13 non-words were then recorded by the researcher on a tape recorder. Care was taken so that no epenthetic vowels were inserted before or between the two phonemes in the double consonant clusters. Five Saudi ESL learners who were postgraduate students at UK universities and whose native dialect was QA were asked to judge if they heard a vowel before or between the first two consonants of each of the 13 non-words. Each subject listened to the non-words individually and one at a time. After each word the researcher asked the subject if he could hear a vowel between the two consonants. For a cluster to be regarded as legal in QA, four of the

---

[44] I call this dialect Qassimi Arabic after the area where it is spoken.

five subjects had to report not hearing a vowel either before or between the two consonants. Only definite answers were counted. This condition applied only to eight clusters. These clusters are /dl/, /mr/, /bw/, /fl/, /kl/, /gr/, /kw/, /θw/. Not hearing epenthetic vowels between the phonemes that were illegal in English (/dl/, /mr/, /bw/) indicates that subjects were primarily using their L1 phonotactic knowledge in judging the legality of these clusters.

The second 12 items of the 36 test items were monosyllabic non-words with high phonotactic probability in English as determined by two measures discussed below. The last 12 items were monosyllabic non-words with low phonotactic probability in English (see Appendices A.3-A.7) for a complete list of the items used in this task). Restricting the stimuli to monosyllables helped the researcher avoid the complexities of stress placement and syllabification (Bailey and Hahn 2001). Unlike the first 12 items, which contained illegal initial consonant clusters, the other 24 items consisted of sequences of phonemes that are perfectly legal in English. They only differed on phonotactic probability. There were also 20 monosyllabic real word fillers which were not included in the analysis. Similar to Bailey and Hahn (2001) it was believed that including real words would encourage the processing of the items as "linguistic entities and not as disembodied sound sequences" (ibid: 574).

The two measures used to determine phonotactic probability were as follows: (a) positional segment frequency (i.e. how often a particular phonetic segment occurs in a position in a word), and (b) biphone frequency (i.e. the segment-to-segment co-occurrence probability). To calculate these two probabilities, the researcher used a web-based interface designed by Vitevitch and Luce (2004). The estimates used in this phonotactic probability calculator were derived based on log-frequency-weighted counts of words in the 1964 *Merriam-Webster Pocket Dictionary* which contains about 20,000 English words, and the log values of the frequencies with which those words occurred in English based on the counts in Kucera and Francis (1967).

In Leigh and Charles-Luce (2002) high and low phonotactic probabilities were computed on 952 CVC words. Based on these words, a median split was determined for the sum of the positional frequency (0.144) and (0.0047) for the sum of biphone frequency. In the current study the same median for the sum of the positional frequency was used. However, the sum of biphone frequency was (0.0045) instead. The reason for this is that one high probability item in the current study had the sum of biphone frequency of (0.0047) and other two had the sum of biphone frequency of (0.0046). The slight change in the median split was not expected to be significant as the highest sum of biphone frequency in the low probability items was (0.0027). Sums above these medians (i.e. (0.144) and (0.0045)) were considered high probability patterns and sums below these were considered low probability patterns.

Non-words that were classified as high probability patterns consisted of segments with high phoneme positional and high biphone probabilities. For example, in the high probability non-word /tæl/, the consonant /t/ is of high frequency in the initial position, the vowel /æ/ is relatively frequent in the medial position, and the consonant /l/ is frequent in the final position in English. A high probability pattern is the sum of these positional frequencies. In addition, the biphone co-occurrences consisted of high probability initial consonant-vowel and vowel-final consonant sequences (e.g., /t/ followed by /æ/ and /æ/ followed by /l/ in the non-word /tæl/. A high probability pattern is the sum of these biphone frequencies. Non-words that were classified as low probability patterns consisted of segments with low segment positional probabilities and low biphone probabilities in American English. For example, the low probability non-word /zuð/ consists of a low probability initial consonant /z/, a medial vowel /u/ and final consonant /ð/. Additionally, the biphone probability of initial consonant-vowel sequences (/zu/) and vowel-final consonant sequences (/uð/) are low frequency co-occurrences. Like the high probability patterns, the low patterns are the sums of the positional frequencies and the biphone co-occurrences. For the high probability phonotactic non-word stimuli, the average segment and biphone probabilities were (.1754) and (.0066), respectively. For the low probability stimuli, the average segment and biphone probabilities were (.0443) and (.0008), respectively. For complete lists of the items used in this task with their probability statistics see Appendix A.3- A7).

Because EFL learners' sensitivity to phonotactic probability in English was to be tested, a very important point had to be taken into consideration in preparing the 24 high and low phonotactic probability non-words. As discussed in Chapter 3, in most if not all studies that tried to examine the sensitivity to phonotactic probability, native speakers' (usually children or monolinguals) sensitivity to the phonotactic probability of their native language was examined. In this case, the researcher may not have to worry about subjects using their phonotactic knowledge of another language in responding to the stimuli. In the current study, the case is different. Asking EFL learners to rate how English-like a non-word is (in the Non-word Rating Task) or to decide if a monosyllabic sequence is a real English word or not in an online task (the Lexical Decision Task, discussed below) does not guarantee that those learners will use only their knowledge of the target language (English) phonotactics. It is not unlikely that subjects in this study will use their sensitivity to the phonotactic probability of their L1 (QA) when the phonemes used in constructing the non-word stimuli are also available in their L1 inventory. This is more likely to happen in the lexical decision task due to the time constraint.

As a result, the researcher had to take two measures in constructing the non-word stimuli for the current study. Recall that phonotactic probability is determined by: (a) positional segment frequency (i.e. how often a particular phonetic segment occurs in a position in a word), and (b) biphone frequency (i.e. the segment-to-segment co-occurrence probability). In CVC sequences, the latter is heavily reliant on the medial vowel (nucleus) because, as discussed above, the biphone frequency is the sum of the initial consonant-vowel and vowel-final consonant frequencies (e.g., /t/ followed by /æ/ and /æ/ followed by /l/ in the non-word /tæl/.

Therefore, the first measure I took to prevent L1 interference was to choose English vowels in the medial position that are missing in both SA and QA (/ʌ/; /aɪ/; /e/; /ɔɪ/). This measure also affected the positional probability which is the sum of the three positional frequencies including the medial segment (vowel). The other measure was to include a consonant in the initial or final position that is also missing in SA and QA (v, p, ʒ). This way, if subjects still use their L1 phonotactics in responding to the stimuli, this will only affect their judgment of the positional frequency of one segment assuming that the positional frequency of this segment is different in the two languages. In four items of the high probability stimuli a different measure was used.

[45] This measure involved using a vowel that is used in SA and QA (/ɪ/ and /æ/) but using two of the missing consonants in SA and QA (i.e. v, p and ʒ) in initial and final position. This measure will lead to the same result attained by using the first two measures.

4.4.1.2 Procedure

The 56 items (20 real word fillers and 36 test items) were randomised and then recorded via a microphone onto a computer. Recordings were done by a female native speaker of North American English using Cool Edit Creative Wave Studio software, version 5.00.06. They were sampled at 22.050 kHz 16 bit, mono. These recordings were then edited into a sound file, saved on a computer hard disk and saved on disc. The native speaker was asked to avoid inserting epenthetic vowels before or between consonants in the items starting with illegal English consonant clusters.

Since the effect of orthography on our results by using spelling or phonetic transcription of the items was possible, a better way to help subjects track the items was used. The native speaker recorded the numbers from 1-56 onto the same computer software and the numbers were then saved in another sound file. After that, the numbers were inserted before each of the 56 items producing a sound file with numbered items. The between-item silence was next adjusted so that there was exactly a four-second silence between items. This sound file was then recorded into a tape recorder using a digital-analogue converter.

---

[45] This was due to inability to create more high phonotactic probability items using the first two measures.

Whereas native speakers were tested individually using this and all other tasks, EFL subjects were pre-tested with this task in one group. Native subjects performed the Non-word Rating Task individually because first, they were given different convenient time slots to perform the task and second, a language lab was not easily accessible even if they had been able to be tested at the same time. However, all subjects performed this task after performing the Lexical Decision one. There was a good reason for this. Since the same test items were used in the Lexical Decision Task, piloting as discussed below, showed that a second presentation of the items in the Lexical Decision Task affected reaction times (henceforth RTs) more than a second presentation of the items in the Non-Word Rating task affected ratings.

When testing native speakers and to ensure maximum understanding of the task, the researcher thoroughly explained to each native subject the requirements of the task. Subjects were first given a questionnaire (see appendix A.8 numbered from 1-56 with a scale of typicality from 1-5 and RW standing for real word at the right of the highest score in the rating scale under each number as in (2).

(2) Very non-typical = 1 2 3 4 5 RW= Very typical

Each subject was told that s/he would listen to a list of 56 numbered items. Some of these items were real English words and others were made up words. The subject's task was to judge how English-like each item sounded and to record their judgment by circling a number on the scale with number (1) representing the least English-like and number (5) the most English-like. If the subject thought that the item was a real word s/he had to circle RW standing for 'real word' which was the sixth option on the scale.

Testing their subjects with the word likeness (Non-word Rating) task, Frisch et al. (2000) provided them with descriptors for each point on the 7 point scale they used. A rating of 1 for example was described as "low-impossible-this word could never be a word of English". In the present study, it was believed that providing subjects with such a clear description of what each number on the rating scale means was an indirect way of guiding them to choose what they were expected to. Recall that some of our items were low probability and others were low probability rimes preceded by illegal clusters. Therefore, telling subjects that number (1) means that the non-word can never be a word in English could have guided them to assign number (1) to non-words starting with the illegal cluster all the time. In the current study the aim was to know how subjects' implicit knowledge, not explicit knowledge and an explicit description of what each rating means, would lead them to differentiate between the two categories.

Using headphones, native English-speaking subjects listened individually to the aurally presented items at a comfortable listening rate with 4 second-intervals between the items. Subjects first listened to three practice items to familiarize them with the task. These were one high phonotactic probability item, one low phonotactic probability item and an item starting with illegal consonant cluster. For the non-native subjects the same procedure was followed except that the subjects were tested in one group and the items were presented using an audio tape played using a Sony analogue language laboratory. Non-native subjects sat in separate booths and listened using built-in headphones. For a complete list of the items in the order they appeared in the task see appendix A.9.

4.4.1.3 Research questions and hypotheses

The main research question that the Non-word Rating Task tried to answer was if Arabic speaking EFL learners are sensitive to both the legality and probability of English phonotactics and the legality of Arabic phonotactics when presented with EFL items. More specifically, this task tried to answer our research question (1) Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English and Arabic? And research question (2) Are Arabic speaking EFL learners sensitive to the phonotactic probability of English? However, since a native English-speaking control group took part in this study, I will present the hypotheses regarding the native control group's results first. For the native group, it has been hypothesised that they will show sensitivity to both constraints and probability of English phonotactics and therefore the relevant hypotheses are as follows:

Hypothesis 1: Native English speakers will rate the stimuli items in the *illegal in 2* (English and Arabic) and *illegal in 1* (English) conditions lower and with fewer errors (classifying a non-word as a real word) than those in the *low probability* condition.

Hypothesis 2: Native English speakers will rate the stimuli items in the *low probability* condition lower and with fewer errors than those in the *high probability* condition.

Certain facts encourage predicting a particular performance by the non-native subjects in the Non-word Rating Task. Recall that our subjects are low-intermediate level EFL learners of English. Therefore, their chances of having been exposed to much naturalistic English input, which would have built their sensitivity to the phonotactic probability of English, are minimal (see Table 4.1). On the other hand, and in addition to the above reason, our non-native subjects were not taught English phonotactic rules,[46] and therefore one may assume that they would not show sensitivity to phonotactic legality in English. However, all this information is hypothetical and therefore null hypotheses were taken regarding non-native subjects' performance in the Non-word Rating Task. These are as follows:

Hypothesis 3: Non-native speakers will show no difference between their ratings or error rates in the *illegal in 2* condition, *illegal in 1* condition and the *low probability* condition.

Hypothesis 4: Non-native speakers will show no difference between their ratings or error rates in the *high probability* condition and the *low probability* condition.

4.4.2 Auditory Lexical Decision

The Lexical Decision Task has previously been used in two different forms: the visual form (e.g. Altenberg and Cairns (1983)) and the auditory form. The latter is the one used in the current study. This task requires subjects to classify aurally presented items into words and non-words and reject non-word items as fast as possible by pressing a computer key, which then measures reaction time (see section 4.4.2.2 for a more detailed description of the procedures). This task has been used in many psycholinguistic studies to test the effect of variables such as phonotactic probability and neighborhood density on lexical access (see Vitevitch and Luce 1999). Since this task requires the correct classification of items into words and non-words, it has been argued that this task entails some lexical processing (Goldinger 1996) in order to arrive at the correct classification. Consequently, the speed of classification (i.e. reaction times in this task) is dependent on the phonological nature of items.

---

[46] This was revealed by the survey conducted in the first week of the treatment.

In the current study, the main concern was with the legal non-words that differed in phonotactic probability and neighborhood density and the illegal ones. Based on the nature of the Lexical Decision Task, it has been assumed that there should be a difference in the speed of classifying these items as non-words. Non-words that are high in phonotactic probability will be responded to more slowly than those with low phonotactic probability. On the other hand, classifying non-words with illegal initial consonant clusters will be the fastest because it is assumed that these items will be ruled out at an earlier stage because of their early isolation point (see Chapter 3 for discussion).

The other motivation for using the Lexical Decision Task in the current study is to find out what type of knowledge subjects are using. The rating task may give subjects a chance to use their metalinguistic knowledge in performing the task. Although subjects may not be able to use their metalinguistic knowledge in judging the phonotactic *probability* in the rating task, it is very likely that they will use this knowledge provided by explicit teaching in judging the phonotactic *legality* of the illegal initial consonant clusters in the same task. Therefore, The Lexical Decision Task aims at finding out if subjects have acquired native-like knowledge which would allow them to respond automatically to illegal consonant clusters starting illegal non-words. It also aims to examine in more detail subjects' sensitivity to the phonotactic probability attained in the Non-Word Rating task.

4.4.2.1 Materials for the Lexical Decision Task

The same 36 test items used in the rating task were used in this task. There were also 24 monosyllabic real word fillers. These were the same 20 items used in the rating task in addition to four new ones.[47] (See Appendix A.10). In addition to the main variable (phonotactic probability and legality) three other sets of statistics were computed for the non-word stimuli to insure that all potential confounding variables were controlled. These variables are as follows:

1. Neighbourhood Density: this was computed by comparing a given phonemic transcription (constituting the stimulus) to all other transcriptions in an on-line version of *Webster's Pocket Dictionary*, mentioned above in Section 4.4.1.1. A neighbour was defined as any transcription that could be converted to another word by any one-phoneme substitution, deletion, or addition. This condition produced a number of these. The statistics were then computed with the help of Vitevitch (personal communication 31.10.2005).[48]
2. Isolation Point: Luce and Large (2001: 569) define the isolation point as "the point at which each word or non-word stimulus diverges from all possible words in the lexicon, yielding either a single possible word (uniqueness point), or no possible words (non-word point) respectively". Controlling the isolation points in our stimuli was crucial. [49] Without doing so, any difference in the reaction times could be attributed to a difference in the position of the isolation point rather than a difference in phonotactic probability. The isolation point in the *high probability* and *low probability* stimuli was always

---

[47] Since the lexical decision task requires the classification of items into words and non-words increasing the number of real words was helpful in keeping subjects focused.

[48] I emailed my stimuli items to Vitevitch and he kindly calculated their neighborhood density using the procedure mentioned above and emailed me the statistics for each item back.

[49] This did not apply to the stimuli starting with illegal consonant clusters where the isolation point was always on the second phoneme.

on the third phoneme except in two items in the *low probability* condition, where it was on the second one. This statistic was computed by the researcher using *Oxford Advanced Learner Dictionary.*

3. Stimuli Duration: It has been argued (Goldinger 1996) that difference in mean duration of the stimuli in different conditions can either reduce or inflate the effect that the condition is expected to produce. In our case if the mean duration of the stimuli in the *low probability* condition, for example, was significantly longer than that in the *high probability* condition, this would have consequently produced longer RTs measured from stimulus onset in the *low probability* condition. Therefore, mean duration had to be measured to ensure that such confounds did not exist. Using Cool Edit Creative Wave Studio software, version 5.00.06 the silence before and after the actual stimuli was removed and then the durations of the sound files were measured. This procedure is very important because in some studies (Vitevitch and Luce 1998; 1999) researchers, as they later conceded (Vitevitch and Luce 2005), erred by reporting the duration of the entire sound files (including leading and trailing silences as well as the stimulus itself) instead of the durations for just the stimuli. Such error could have confounded the present results.

The five statistics including phoneme positional and biphone probabilities are shown in Table 4.2.

Table 4.2 Mean scores for the five statistics performed on the stimuli in four conditions

Condition	Average Phoneme position probability	Average biphone probability	Average Density	Average Isolation point (phonemes)	Average Duration (ms)
illegal in 2	.0711	.00125	.833	2.00	710
illegal in 1	.0655	.00151	.500	2.00	698
high probability	.175	.00660	19.66	3.00	723
low Probability	.044	.000841	4.250	2.833	775

A one-way ANOVA on each of the five scores with the four conditions as independent variables was performed. This was first to confirm expected significant differences (i.e. average phoneme and biphone probabilities and average density between high and low probability conditions), and second to ensure that potential confounds (i.e. differences in average isolation points and durations) were controlled. The results of post hoc test showed that the items in the *high probability* condition had a significantly higher ($p<.005$) phoneme position probability, biphone probability and neighbour density than any of the other three conditions.

The items in the *low probability* condition also had a significantly ($p<.05$) lower phoneme position probability than the items corresponding to the syllable starting at the second consonant of the non-word items in the *illegal in 2* and *illegal in 1* conditions. However, it should be noted that this difference works against our hypothesis regarding the effect of phonotactics. Initial illegal clusters are associated with faster RTs, whereas a higher phoneme position probability is one of two factors regarding phonotactic probabilities that are associated with slower RTs in the Lexical Decision Task. Thus if items starting with illegal clusters were still judged less English like than those with low phonotactic probability, this would show the strong effect of the illegal initial clusters.

Except for the predicted earlier isolation points that the stimuli in the *illegal in 2* and *illegal in 1* conditions had than those in *low* and *high probability* conditions (significant at $p<.005$), the analysis showed that potential confounds were well controlled. There were no other significant differences ($p<.05$) between conditions, neither in isolation points nor for durations.

4.4.2.2 Procedure

The 60 items were recorded one at a time in a list by the same native speaker using Cool Edit Creative Wave Studio software, version 5.00.06. They were sampled at 22.050 kHz 16 bit, mono. Each stimulus was then edited into an individual sound file, saved on a computer hard disk and saved on disc. The items were then transferred to the hard disk of a portable computer. Using the same software, silence at the beginning and the end of each file was then removed leaving only the actual stimulus.

In this task, all participants (this time including non-native subjects) were tested individually. Presentation of the stimuli and measuring reaction times was controlled by a portable computer running DMDX experiment control software which randomised the stimuli for each subject. Each subject listened to the stimuli over headphones. Subjects were told that they would hear a list containing both real words and made-up words. A "go/no-go" procedure was used in this task. In other words, when presented with one of the stimuli items over headphones at a comfortable listening level the subject had to respond by pressing the mouse key[50] using her/his dominant hand as quickly and accurately as possible when s/he heard a non-English word. For no-go, when the subject heard an English word s/he did not have to do anything but wait for the next item. The computer measured and stored the reaction times. In this task, response latencies were measured from the onset of each stimulus

---

[50] Since a laptop was used, the mouse key represented a more comfortable option than a keyboard key for both left-handed and right-handed subjects.

file (i.e. item) to the onset of the participant's response. Recall that silence at the beginning of the sound file was removed. Therefore, the onset of the sound file coincided with the onset of the actual stimuli items. After registering a response, the computer began the next trial and the subject heard the next item. Participants were allowed a maximum of three seconds to respond before the computer automatically recorded a null response and presented the next item.

Prior to the experimental items, each subject received three trial items (see appendix A.11). These were two non-words from different phonotactic probability conditions and a real word. None of the items used in the practice session were used in the experiment. These items were used to familiarize the participants with the task, and the data collected from them were not included in the final analysis.

4.4.2.3 Research questions and hypotheses

This task tried to answer the same research questions (1 and 2) as in the Non-word Rating Task: (1) Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English and Arabic? And research question (2) Are Arabic speaking EFL learners sensitive to the phonotactic probability of English? For the native group, it has been hypothesised that they will show sensitivity to both constraints and probability of English phonotactics and therefore the relevant hypotheses are as follows:

Hypothesis 5: Native English speakers will reject stimuli items in the *illegal in 2* and *illegal in 1* conditions faster and with fewer errors than those in the *low probability* condition.

Hypothesis 6: Native English speakers will reject stimuli items in the *low probability* condition faster and with fewer errors than those in the *high probability* condition.

Hypothesis 7: Native speakers will be faster in rejecting stimuli items and with fewer errors in all conditions than non-native speakers.

Because of the quick processing the Lexical Decision Task requires, it was predicted that there will be a difference between the performance of the native and the non-native speakers regarding their sensitivity to phonotactic probability of English. More specifically, it was assumed that non-native speakers will not show a difference in their RTs to the stimuli items in the low probability and high probability conditions because of the fast processing that would be required. However, because of their limited lexicon, they are likely to assume that high probability items are real words and therefore a difference in the error rate is expected.

On the other hand, it was predicted that non-native speakers would show sensitivity to the phonotatic legality represented in the stimuli items in the *illegal in 2* and *illegal in 1* conditions both in RTs and error rate. That is because, unlike the situation in the *high probability* and *low probability* conditions in which the subject has to process the whole sequence to give a judgement, the stimulus items in the *illegal in 2* and *illegal in 1* conditions can be rejected based on the first two phonemes (i.e. isolation point). Having said this one can not rely on such predictions in formulating hypotheses regarding non-native subjects' performance in the Lexical Decision Task. That is because of the burden that this task is expected to have on subjects' processing of EFL words. Thus, null hypotheses were taken regarding non-native subjects' performance in the Lexical Decision Task. These are as follows:

Hypothesis 8: Non-native speakers will show no difference in RTs or error rates in rejecting stimuli items in the *illegal in 2* condition, *illegal in 1* condition and *low probability* condition.

Hypothesis 9: Non-native speakers will show no difference in RTs or error rates in rejecting stimuli items in the *high probability* condition and *low probability* condition.

4.4.3 Word Spotting Task

In the current study, the Word Spotting Task was used to answer our third research question, namely, do Arabic speaking EFL learners' use the phonotactic constraints of English and Arabic in lexical segmentation of running speech in English? The Word Spotting Task basically requires subjects to spot real words embedded at the beginning or the end of nonsense sequences. Unlike some other psycholinguistic tasks (e.g. Lexical Decision) which entail procedures that are not part of natural language processing, it has been noted (McQueen 1996) that word spotting has ecological validity. That is, although the sequences in the task are shorter, it resembles the listener's normal task of spotting words in continuous speech. This task has proven to be effective in investigating the role that different cues play in lexical segmentation. As discussed in Chapter 3, this task has shown that metrical structure (Cutler and Norris 1988; Cutler, Mehler, Norris and Segui 1992), phonotactics (McQueen 1998; van der Lugt 2001; Weber 2000; Weber and Cutler 2006), and allophonic variations (Dumay, Content, and Frauenfelder, 1999) are used to direct segmentation.

Another advantage of this task, as recently argued by Weber and Cutler (2006), is that it allows comparing the detection of the same word in different contexts. This is important because it means that the researcher does not need to control certain properties of target words such as frequency and neighbourhood density, both of which are difficult to determine in relation to L2 listeners with linguistic experience different from L1 listeners'. Producing the required boundary conditions by manipulating the preceding context of the same target word makes it possible to avoid potential confounds.

In using the same item in different contexts we can safely attribute the results relating to dependent variables to different context and not to the nature of the item per se. With that in mind, 36 English words that could appear in different contexts were selected. Description of the design, selection process, and the criteria used are presented in the next section.[51]

4.4.3.1 Materials for the Word Spotting Task

The aim of this task was to find out if Arabic-speaking learners of EFL use the phonotactic constraints of their L1 (QA) and FL (English) in the segmentation of continuous EFL speech. English words whose initial phoneme could be manipulated to produce four preceding contexts were therefore required. In one condition, the word boundary is signalled in both languages. In a second condition, the boundary is signalled only in English. In a third one a boundary is signalled only in Arabic and in the last condition no boundary was signalled in either language.

Recall that using a certain procedure we could identify only eight initial clusters (/dl/, /bw/, /mr/, /fl/, /kl/, /gr/, /kw/, /θw/) that were legal in QA. This constrained the

---

[51] The design of the Word Spotting Task in this study, particularly the use of four different lists, was inspired by the design of the same task used by Weber and Cutler (2006). However, some changes were made to our task. These will be emphasised where appropriate.

choices that could be made because in the fourth condition (i.e. *No Boundary* condition) words had to start only with the phonemes /l/, /w/ or /r/ so that no boundary in QA was signalled. Luckily, using words starting with these phonemes, from the inventory provided by Yavaş (2006) of illegal English two-member onsets and codas, it was possible to meet the first two conditions (i.e. *Common Boundary* and *English Boundary*) as well. Forming stimuli for the third condition (*Arabic boundary*) was relatively simple. That was because any preceding phoneme other than /d/, /f/ and /k/ before words starting with /l/, phonemes other than /m/ and /g/ before words starting with /r/ and finally any phonemes other than /b/, /k/ and /θ/ before words starting with /w/ could form the required condition.

Recall that in Chapter 3 I argued that consonant clusters that are illegal only in English onsets (e.g. /gd/) may not be decisive for determining word boundaries. In other words, although this cue suggests that the sequence *bigdog* could be segmented as *big dog*, it does not rule out a segmentation like *bigd og* as /gd/ is legal in coda in English, and VC syllables are also allowed in English. Therefore, care was taken in choosing clusters which met our first two conditions. Specifically, for a cluster to provide a boundary in these two conditions it had to be illegal in both the onset and the coda of English. Therefore, a preceding phoneme like /r/ was not used before words starting with /l/ because, whereas the cluster /rl/ is illegal in the onset it is legal in the coda position in North American English (e.g. girl).

With the above considerations in mind, 36 common monosyllabic English words starting with /w/, /l/ and /r/ (12 words each) were selected, using the *Oxford Advanced Learner's Dictionary*. These words appeared in the final position of bisyllabic nonsense strings (such as *loot* in /zautlu:t/). The onset of the embedded target word was either clearly aligned with a syllable boundary or not based on the phonotactics of English and QA. Four alignment conditions were used in this task. In one condition, both languages require a syllable boundary at the onset of the word *line* as in /vi:ʃlain/. Words in both languages can not start with /ʃl/. In another condition, only English requires a syllable boundary as in /vi:dlain/. Words in QA but not in English can start with the consonant cluster /dl/. In a third condition, only QA requires a syllable boundary as in /vi:blain/. English but not QA words can start with the cluster /bl/. In the last condition, neither language requires a syllable boundary as in /vi:flain/. /fl/ is a possible syllable onset in both languages.

In creating the nonsense syllables, where possible no real words in either language were embedded in these syllables. Therefore, a syllable like /kauk/ which is not a word in either language was not used because /kau/ *cow* is a real English word. In addition, where possible, target words were carefully chosen so that when English does not require a syllable boundary, the final phoneme of the nonsense syllable could not create a word other than the target word (e.g. the target word *lick* in the nonsense sequence /zɔɪklɪk/ could be realized as *click*).

In addition and similarly to Weber and Cutler (2006), the vowels used in the preceding nonsense syllables were either long or diphthongs. This way the syllable was phonotactically legal in English without the coda and therefore "the internal structure of the context syllable did not itself force a particular segmentation" (ibid: 599). Thus, the two segmentations (e.g. /zaɪ#fli:f/ and /zaɪf#li:f/) are both legal in English.

Recall that Weber and Cutler (2006) used different nonsense syllables before the same target word in the four different conditions (e.g. *fumloft*, *prarloft*, *forsloft*, *zarploft*). A better way would have been to keep the preceding context the same for all conditions of the same word and only change the last phoneme which determined the alignment condition. This way they could have minimised the effect that the use of completely different nonsense syllable could have on the pronunciation and recognition of the target word. Therefore, unlike Weber and Cutler, the nonsense syllable preceding each word in the current study was identical in the four conditions except for the final phoneme which determined the alignment condition. As I said, this was to control for the effect that the structure of the preceding nonsense syllable as a whole may have on the recognition process of the target word. A completely different preceding context could affect the natural pronunciation of the following target word, thus producing a potential confound. However, different nonsense syllables were used before other target words starting with the same initial phoneme.

Four lists of items were made. Each list contained the 36 target-bearing sequences (with each word appearing only once in a particular alignment condition and). The four types of boundary condition were counterbalanced across lists. Each list also contained 24 fillers. These fillers were 10 bisyllabic nonsense sequences which contained monosyllabic words in final position starting with phonemes other than /l/, /r/, and /w/ and 14 bisyllabic nonsense sequences which contained no English or Arabic words. The structure of the initial nonsense syllable in the filler items was identical to those used before the target words. There was also three practice items which appeared at the start of each list. There was another advantage of using four different lists, namely ruling out the effect that the repeated presentation of the same target word would have on results. All items used in the Word Spotting Task including the practice items and fillers are in appendix B.1-6.

4.4.3.2 Procedure

The 168 bisyllabic sequences (144 target-bearing and 24 fillers) were recorded by the same female native speaker of North American English using Cool Edit Creative Wave Studio software, version 5.00.06. They were sampled at 22.050 kHz 16 bit, mono. The speaker was asked to avoid any clear syllable boundaries in her production by not pausing when there was an English boundary. Using the same software, silence at the end of each of the 36 target-bearing files was removed. Each stimulus was then edited into an individual sound file, saved on a computer hard disk and saved on disc. These files were then transferred to the hard disk of a portable computer.

Presentation of the stimuli and measurement of RTs was controlled by a portable computer running DMDX experiment control software. Each subject, native and non-native, was tested individually using headphones. They were instructed to listen to the nonsense sequence and press the mouse key using their dominant hand as fast as possible when they detected an embedded English word at the end of the sequence and then say the word aloud. Each subject listened to one of the four lists that were randomised in each presentation. An equal number of subjects in each group (i.e. three native and five in each of the non-native groups) listened to each of the four lists. The computer measured and stored the RTs and oral responses were tape recorded for error analysis. The script of DMDX was adjusted so as to measure RTs from the offset of the sound file which coincided with the offset of the target word.

Prior to the experimental items, each subject received three trial items. These were two bisyllabic real word-bearing sequences and one bisyllabic sequence with no embedded word. None of the items used in the practice session were used in the experimental lists. These items were used to familiarize the participants with the task, and the data collected from them, as was the case with all 24 fillers, were not included in the final analysis.

Although the target words used in this task were very common words, a measure was nonetheless taken to ensure that our non-native subjects actually knew these words. This was to ensure that error rates in this task would reflect subjects' inability to spot words in certain conditions rather than the fact that they did not know these words. Therefore, at the beginning of the semester and one week before taking this task subjects were given a list (see appendix C.1) containing the target words and were asked to provide the phonetic transcriptions of these words.[52] Leaving one week after this procedure and before testing subjects with the Word Spotting Task was necessary because a short lapse of time could have given the non-native subjects an advantage over native subjects by providing them with previous exposure to the target words.

4.4.3.3 Research questions and hypotheses

This task investigated research question 3, namely "do Arabic speaking EFL learners' use the phonotactic constraints of English and Arabic in lexical segmentation of running speech in English?" Previous research indicates that the Word Spotting Task is not an easy one. Davis (2000) has for example noted that even native speakers'

---

[52] Students at this Department are taught the IPA at level 1.

error rates can be as high as 70%. This means that this task requires higher proficiency in the language in question. Therefore, it was predicted that our native speakers will outperform the non-native speakers in spotting words in all conditions. On the other hand, as discussed above, this task has been shown to be sensitive to native phonotactic cues. Consequently, it was predicted that our native subjects would show sensitivity to the phonotactic cues in the current task. However, since our native subjects have no knowledge of Arabic, it is assumed that there should not be any effect of Arabic phonotactic cues on their performance. Therefore, the hypotheses for native speakers' are as follows:

Hypothesis 10: Native English speakers will spot words in the *Common Boundary* and *English Boundary* conditions faster and with fewer errors than those in the *No Boundary* conditions.

Hypothesis 11: Native English speakers will show no difference in spotting words in the *Arabic Boundary* and *No Boundary* conditions.

Hypothesis 12: Native English speakers will be faster and more accurate than non-native speakers in spotting words in all conditions.

Again, and as is the case with the two previous tasks, a null hypothesis was taken regarding non-native subjects' performance in the Word Spotting task.

Hypothesis 13: Non-native speakers will show no difference in RTs or error rates in spotting words in the *No Boundary* condition on the one hand and the other three conditions on the other.

4.4.4 Piloting

Prior to collecting data from our native subjects using the three tasks discussed, these tasks were piloted with four subjects. Two of these subjects were native speakers of British English. The other two were advanced Arabic-speaking ESL learners studying at Newcastle University. One aim of the pilot was to ensure that conducting the experiment using the portable computer would run smoothly and that subjects would be comfortable with certain procedures followed (e.g. intervals between items, etc.). The other aim was to ensure that particularly native subjects would not spot target words in the Word Spotting Task before their offsets. Recall that RTs in this task were measured from the word offset and if subjects spotted the words before their offsets, RTs will not be scored. However, the pilot showed that none of the native subjects could spot a target word before its offset. The fastest RT scored by a native subject was 289 ms measured from offset.

Two of the subjects, one native and another non-native, took the Lexical Decision Task before the Non Word Rating Task and the other two started with the Non-Word Rating Task first. The latter two subjects reported that this sequence (i.e. starting with the Rating task) facilitated identifying non-words in the Lexical Decision Task. But, the other two subjects reported that listening to the non-words in the Lexical Decision Task did not influence their ratings and that they felt inclined to give the same ratings even if they listened to the non-words for a third time. Based on these statements, all subjects included in the study took the Lexical Decision Task first.

Three amendments were suggested by these subjects. These amendments were accepted and applied to the final format of the tasks. These amendments were as follows:

1. The scale used in the Rating Task included five numbers to be chosen. Initially when a real word was heard, the subject had to choose five (i.e. the highest rating). A native subject suggested including a separate category for real

words so that real words were not equated with high probability non-words. This was done by adding the abbreviation RW standing for real word at the far right side of the scale, as is already shown above in Section 4.4.1.2.

2. Based on the feedback from the four subjects, convenient time-intervals between items were chosen in each task. These were the same for native and non-native subjects.
3. Some items with very high error rates were discarded.

4.4.5 Treatment

The actual length of the term when the experiment was conducted was 13 weeks. However, the treatment lasted only 8 weeks. That was because the first week was devoted to getting the information from subjects on which the process of dividing them into experimental and control groups was based. The next two weeks were devoted to pre-tests. Recall that all the non-native subjects took the Rating Task at the same time. They were first tested individually in the Lexical Decision and the Word Spotting Tasks. Testing each of the 40 subjects took between 25-30 minutes. The last 2 weeks were allotted to post-tests.

Originally, there were three (one separate and two consecutive) hours allotted to the listening module at the Level 3 every week. However, after dividing subjects into two groups, an arrangement was made to ensure that subjects in each group received an equal number of teaching hours. Each group took the separate hour and the two consecutive hours in alternate weeks. In other words, the group taking the one-hour session in a week took the two-hour session the week after and vice versa. So, on average, each group had one and a half hours of listening instruction per week.

Top-Up Listening 3 was chosen to be the textbook used in teaching both groups. This textbook focuses on teaching phonological phenomena that characterise connected speech. At the end of each unit, there is a section called *Listening Clinic* in which a particular phenomenon is highlighted. These phenomena included reduction, contraction, assimilation, intrusive sounds, stress and intonation.

For the control group, a unit was presented every week and the relevant phenomenon was thoroughly discussed. Students were then set a weekly task. The task was to choose the conversation or speech of a native speaker of North American English using the radio, TV or the internet and record about 40 minutes of it. The subject then had to transcribe 500 words of the material highlighting the phenomenon that was discussed. Students in the experimental group had to do the same in addition to highlighting the English phonotactic constraints they were taught.

Teaching English phonotactics for the experimental group in this study was in three stages. These were as follows:

Stage 1 (first two weeks) Awareness Raising: The aim of this stage was to explain to students that in English, as is the case with all other languages, there are some constraints on what phonemes can appear together in a syllable. Four main points were presented at this stage. Students were instructed

- that the syllable has a certain structure containing an onset, nucleus and a coda.
- that some phonotactic restrictions (i.e. illegal clusters) apply only to the onset and others apply only to the coda.

- that some constraints apply to both the onset and the coda.
- that native speakers use these as segmentation cues and that the course aim was to help them do the same.

At this stage every student in the experimental group received a list of 12 illegal consonant clusters that could appear neither in the onset nor in the coda of an English syllable. Some clusters were also illegal in QA. However, whereas those illegal in English only (/dl/, /mr/, /bw/) were used as the *English Boundary* condition in the Word Spotting Task, those illegal in both were not used to form the *Common Boundary* condition. Providing students with a large number of clusters would have distracted them and in using a small number, effects could be generalized. This list (see Appendix C.2) was collected based on the inventory of illegal onset and coda clusters in Yavaş (2006). Students were instructed to memorise this list.

Stage 2 (week 3 onwards) Observation: As discussed above, subjects in both groups were set a weekly task in which they had to record conversation and then highlight the phonological phenomenon discussed every week. In addition to this, students in the experimental group had to highlight the illegal English consonant clusters appearing between words (e.g. bad lady). Students were instructed to highlight the clusters that they could hear while listening to material they recorded at home and not while only reading their transcription. That was because after the first task it was noticed that clusters as in *Sarah left* were highlighted indicating that some students were highlighting based on the transcription they made.

Stage 3 (Week 5 onwards) In-class practice: At this stage, students were asked to spot clusters during online processing. That is, unlike the situation in the previous stage in which students could play and replay the material to spot the illegal cluster, at this stage new material[53] was played only once and students had to spot the between-word illegal clusters. The task went as follows. Five students were chosen each time new material was played. Each of the five students had to raise his hand when he spotted an illegal cluster. Once a hand was raised, the teacher (i.e. the researcher) stopped the tape and asked the student to identify the cluster. The tape was then replayed for other clusters to be identified. Once the listening material was finished, another five subjects were chosen and new material was played. Two groups took this activity every week.

It should be noted here that unlike other L2 instruction studies where activities resembled or were identical to the tests, the requirement of this activity was different from that of the Word Spotting Task. In this activity, the student's main concern was with the illegal clusters, whereas in the Word Spotting Task they had to spot the word following the nonsense syllable. Thus, a practice effect in this study was controlled.

There were two main problems observed in the treatment provided. Firstly, there was the problem of assimilation, which was expected. Students reported that in a quite big number of between-word illegal clusters one of the two phonemes in the cluster was completely lost (e.g. bad guy /bægaɪ/). The second problem relates to the in-class activity that was performed. This activity proved very difficult. In the beginning (week 5 of the treatment), students could spot a cluster only after at least the five next words were heard. However, students' performance seemed to improve during the

---

[53] Some of the material used at this stage was from the text. Other material was authentic texts recorded by the researcher.

next several weeks. At week 8, some students were able to spot a cluster after only the next word was presented.

4.4.6 Post-tests

The treatment ceased after week 11, when the remaining two weeks were then devoted to post-tests. Approximately nine weeks separated the administration of the pre-tests and the post-tests, thus reducing the possibility that the outcomes evidenced were the result of memory or practice effects of the pre-tests. The tasks and the same procedures used in the pre-tests were also used in the post-tests.

4.4.6.1 Hypotheses

Recall that our pre-tests were to answer three research questions:

1. Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English and Arabic?
2. Are Arabic speaking EFL learners sensitive to the phonotactic probability of English?
3. Do Arabic speaking EFL learners use the phonotactic constraints of English and Arabic in lexical segmentation of running speech in English?

The treatment aimed at improving only the experimental group's sensitivity to the legality of English phonotactics and their ability to use it in lexical segmentation of English connected speech. On the other hand, both group's sensitivity to phonotactic probability could have naturally improved as a result of their exposure to naturalistic English input through the weekly tasks and the in-class activities that they have been set. However, a null hypothesis was taken regarding improvement towards phonotactic probability.

On the other hand, a null hypothesis has also been taken regarding the improvement that subjects may have in their sensitivity to the phonotactic legality of English and its use in lexical segmentation.

Hypothesis 14: Control and experimental non-native subjects will show no improvement towards their sensitivity to the phonotactic probability and phonotactic legality of English and its use in lexical segmentation at post-test.

However, a particular sequence for any potential improvement is predicted. In other words, it is logically assumed that any improvement should show in subjects' sensitivity to the legality of English phonotactics before an improvement will show in their use of this legality in lexical segmentation. As a result, the relevant hypothesis is as follows:

Hypothesis 15: In the Word Spotting Task, experimental subjects' performance in spotting words in the *English Boundary* condition will improve only if this improvement is supported by a similar improvement in their sensitivity towards non-words starting with illegal English clusters in the Non-word Rating and the Lexical Decision Tasks.

4.7 Summary

In this chapter, a comprehensive description of the methodology of the current study has been provided. Subjects, study design, treatment, and the tasks used have been reviewed. In addition, an account of how these tasks are to answer our research questions was discussed. Null hypotheses were taken regarding non-native speakers' performance. However, relevant hypotheses regarding native speakers' performance in the three tasks were listed. The next chapter presents the results that were obtained in each task in both pre-tests and post-tests along with appropriate discussion.

Chapter 5: Results and Discussion

5.1 Introduction

In this chapter, results of the three groups (i.e. native control, non-native control and non-native experimental) in the three tasks (i.e. Non-Word Rating, Lexical Decision and Word Spotting) will be presented. This chapter is divided into two main sections. Section 5.2 presents the pre-test results and Section 5.3 presents the post-test results. After the results of each task are presented separately under each section, a discussion will summarize results and review how they confirm and accept or disconfirm and reject our set of hypotheses presented in Chapter 4. Section 5.4 will summarise the combined results as they relate to our research questions.

5.2 Pre-test results

5.2.1 Results of the Non-word Rating Task

The three practice items and 20 real-word fillers were not included in the analysis. In this task two dependent variables were analysed, namely informant ratings and error rate. An error was counted when either the subject did not give a rating for a non-word item or wrongly classified the item as a real word. Given the nature of this task with the time it gives for subjects to process the non-word items, the error rate was not analysed as a dependent variable in previous research, as we saw in Chapter 3. However, given the low-intermediate level of our non-native subjects, different conditions in this task are predicted to show effects in terms of errors, hence the need for the analysis of error rate in this task.

Recall that the current task investigates two different research questions, namely subjects' sensitivity to the *legality* of English phonotactics and their sensitivity to the *probability* of English phonotactics. Previous research using the Non-word Rating Task has tried to answer one question or the other but not both. Therefore, in studies which investigated subjects' sensitivity to the phonotactic probability of English, for example, only two conditions were compared: *low phonotactic probability* and *high phonotactic probability*. In the current task, however, four different conditions were compared: *illegal in one* (English only), *illegal in both* (English and Arabic), *low phonotactic probability* and *high phonotactic probability*.

Based on this, the best way to answer our two research questions is to compare ratings and error rates in all conditions to the *low phonotactic probability* condition. To answer the first research question, ratings and error rates in the *illegal in one* condition and *illegal in two* condition will be compared to those in the *low phonotactic probability* condition. If subjects are sensitive to the legality of English phonotactics, they are predicted to rate non-words in the *illegal in one* and *illegal in two* conditions as less English-like and with fewer errors than those in the *low phonotactic probability* condition. To answer the second research question, ratings and error rate in the *high phonotactic probability* condition will also be compared to those in the *low phonotactic probability* condition. If subjects are sensitive to the phonotactic probability of English, they are predicted to rate non-words in the *high phonotactic probability* condition as more English-like and with more errors (as they are similar to more real words) than those in the *low phonotactic probability* condition. In the following discussion, effects are considered statistically significant at $p \leq 0.05$ and marginally significant where $0.05 < p \leq 0.1$.

5.2.1.1 Native Control Group

The first procedure was meant to find out if our native English-speaking control group showed effects of different conditions. Their mean ratings and error rates for each condition are shown in Figure 5.1 and Figure 5.2, respectively.

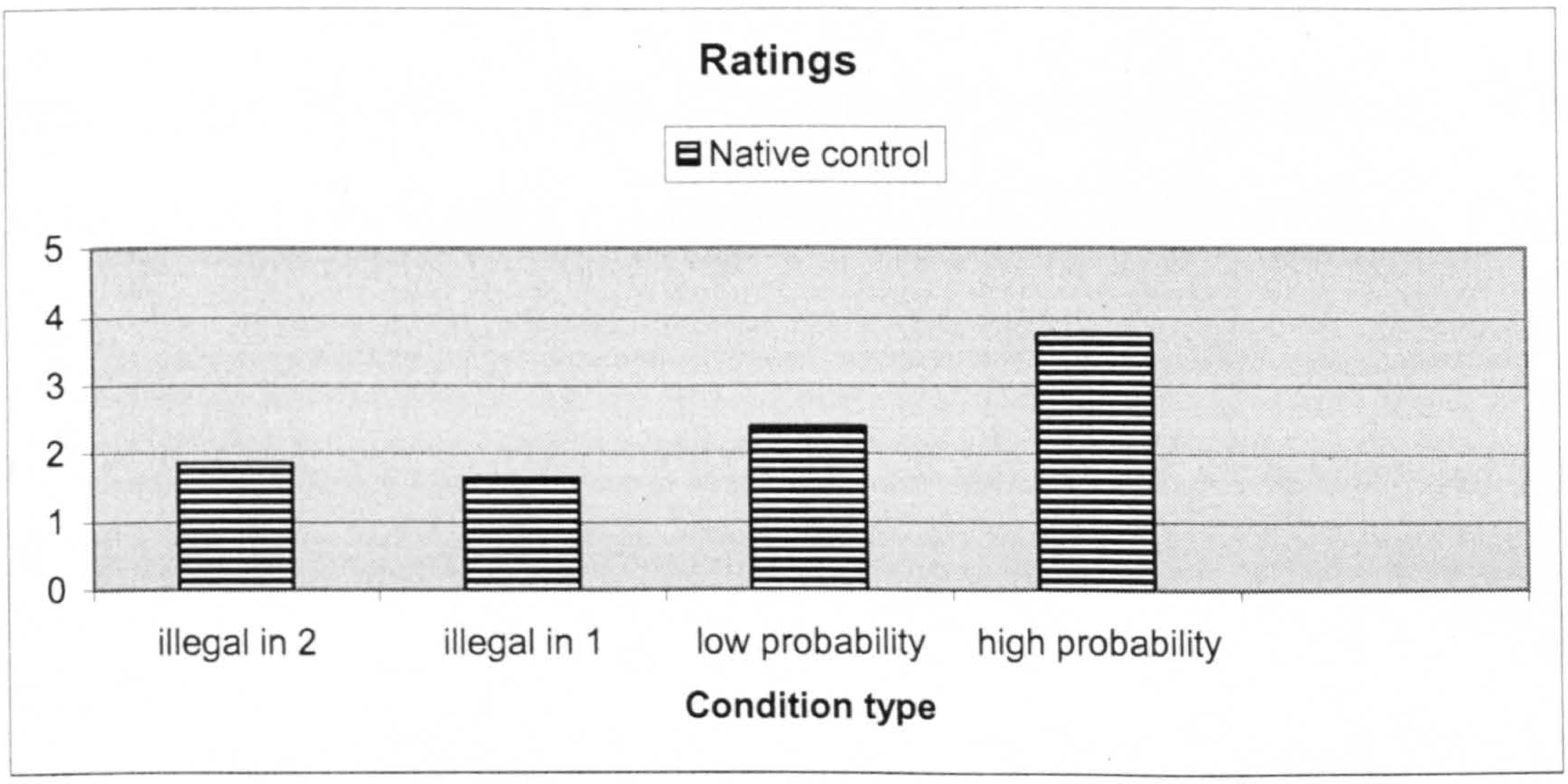

Figure5.1. Native English speakers' out of 5 mean ratings in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2).

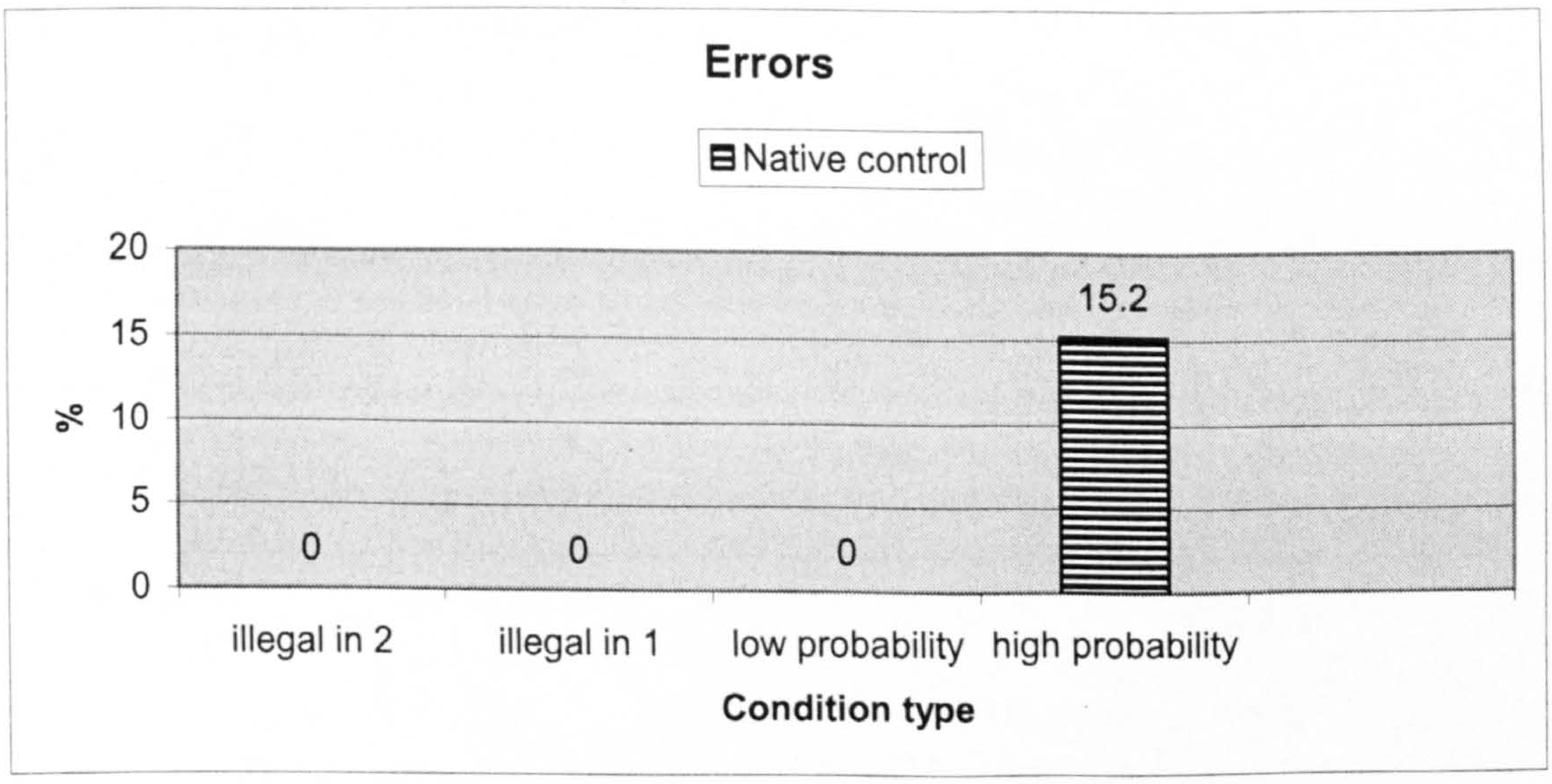

Figure5.2. Native English speakers' mean percentage of errors in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2).

As Figure 5.2 shows, native English speakers did not wrongly judge non-words as real words except in the *high probability* condition. Two separate One-way ANOVAs were conducted to compare native speakers' mean ratings (in the first ANOVA) and error rate (in the second) in the *high probability, illegal in 1* and *illegal in 2* conditions to those in the *low probability condition.* Post hoc test results showed that native speakers' ratings and error rates in the *high probability* condition were significantly higher than their ratings and error rates in the *low probability* condition (Ratings, $p = 0.000$; Errors, $p = 0.000$).

On the other hand, native English speakers had a significantly $p < 0.05$ higher rating but not error rate in the *low probability* condition than those in the *illegal in 2* condition (Ratings, $p = 0.021$; Errors, $p = 1.000$) and *illegal in 1* condition (Ratings, $p = 0.002$; Errors, $p = 1.000$). No significant difference was found neither in the ratings $p = 0.332$ or in the error rate $p = 1.000$ between the *illegal in 2* and the *illegal in 1* conditions.

5.2.1.2 Comparing non-native groups (between groups comparison)

The second procedure in analysing the Non-word Rating Task data aimed to find out if there was any significant difference between the ratings and errors of the two non-native groups (i.e. non-native control and non-native experimental) in the four different phonotactic conditions. Mean ratings and percentage of errors of the two non-native groups in the four conditions are compared in Figure 5.3 and Figure 5.4 respectively.

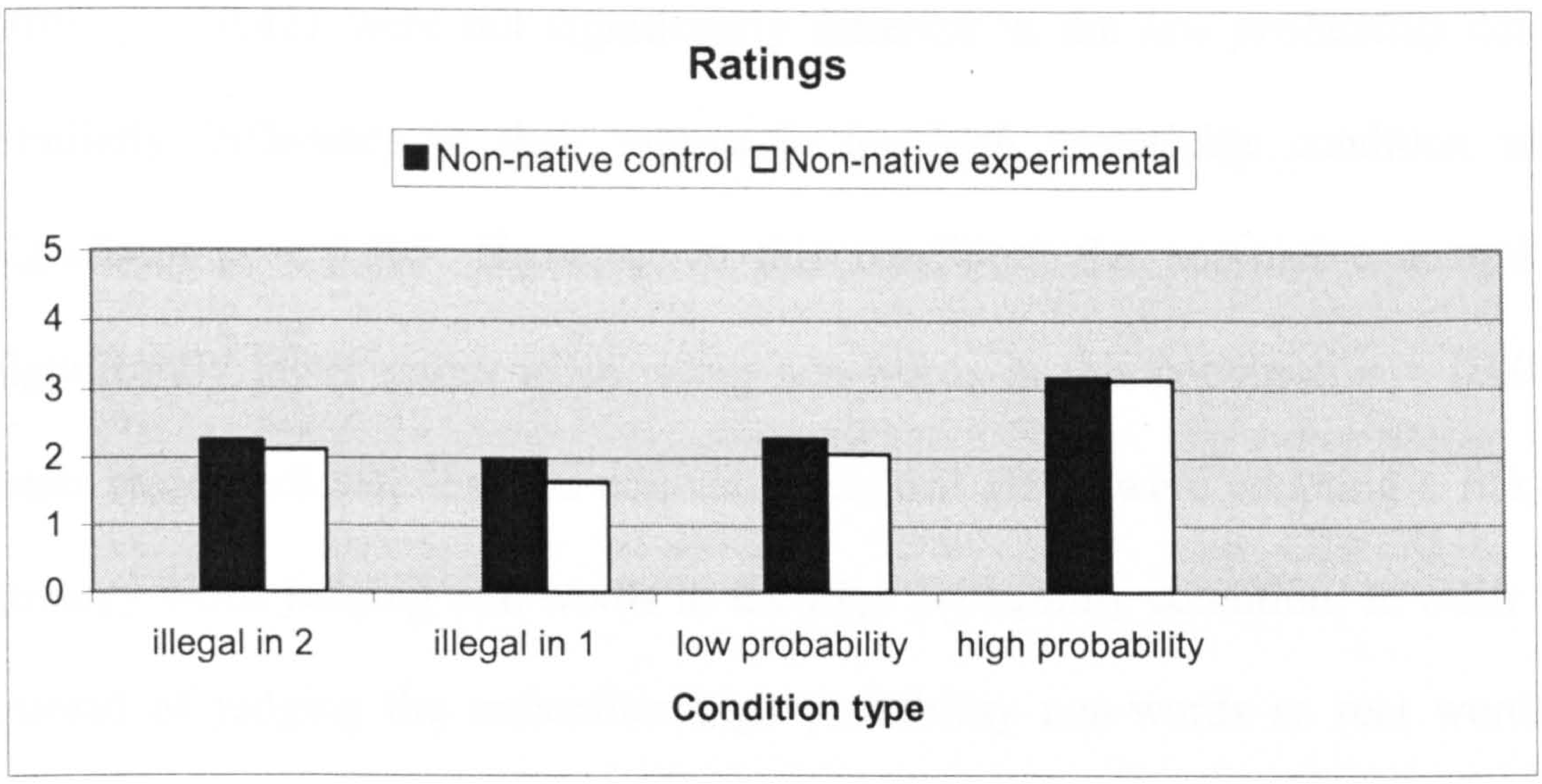

Figure 5.3. Non-native control and non-native experimental groups' ratings in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2).

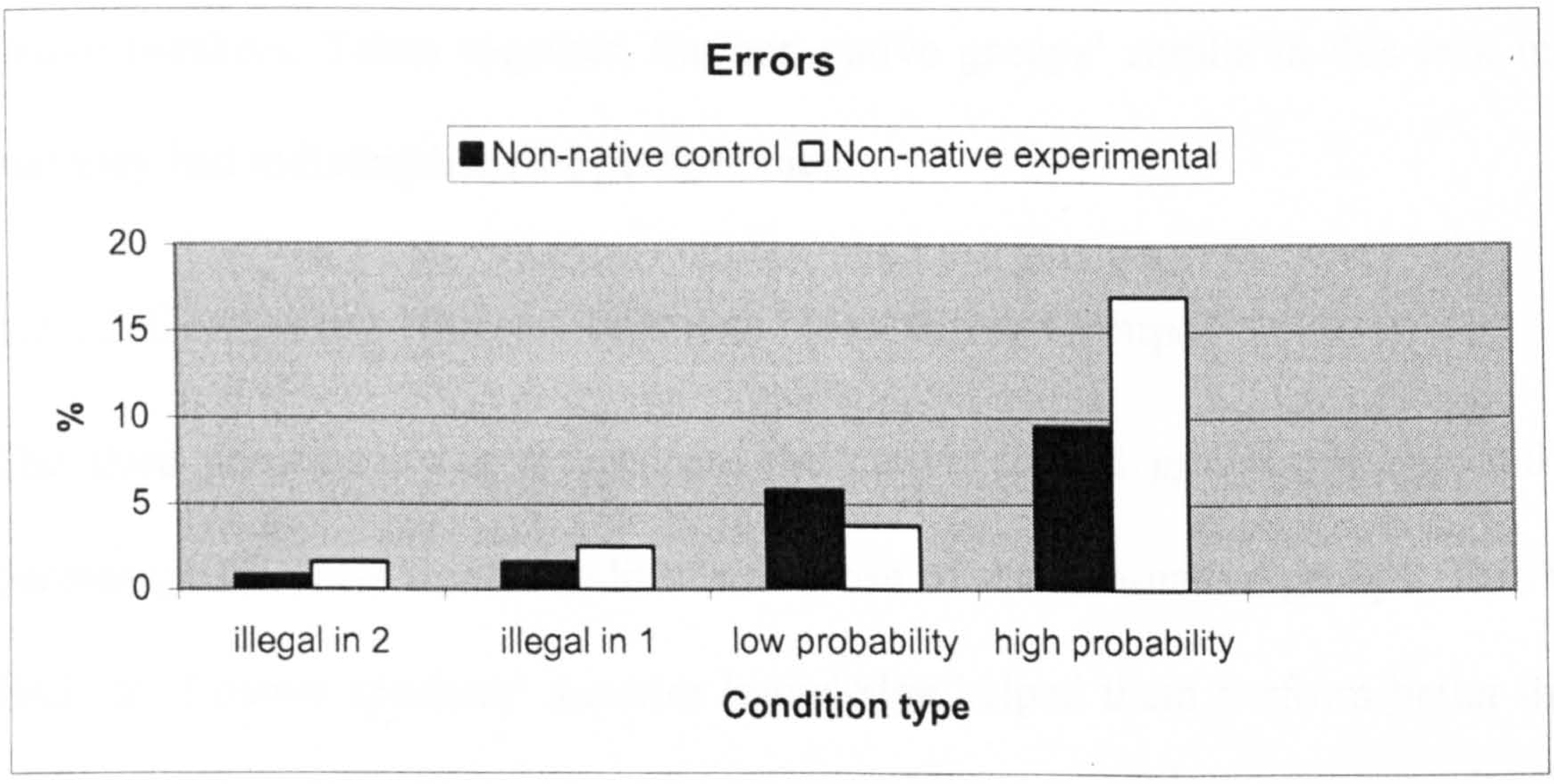

Figure 5.4. Non-native control and non-native experimental groups' mean percentage of errors in the Non-word Rating Task in four different phonotactic conditions (high probability, low probability, illegal in 1 and illegal in 2).

Two separate one-way ANOVAs were conducted to compare mean ratings (in the first ANOVA) and error rate (in the second) of the two non-native groups in all four conditions. Post hoc test results showed that in the *illegal in 2* condition, the two groups' ratings $p = 0.401$ and error rate $p = 0.748$ were not significantly different. Their error rate was also not significantly different in the *illegal in 1* condition $p = 0.748$. However, in this condition, although the difference was just significant $p = 0.044$, the non-native experimental group rated items less English-like than did the non-native control group. Results also showed that both groups' ratings $p = 0.167$ and errors $p = 0.421$ were not significantly different in the *low probability* condition. Similarly, difference in their ratings in the *high probability* condition was not significant $p = 0.785$. However, in this condition, the non-native control made significantly fewer errors when rating non-words in this condition $p = 0.004$. The latter result indicates that the non-native control group were adopting a risk-taking strategy when judging non-words in the *high probability* condition. In other words, instead of judging the unfamiliar high probability non-words as real words, they preferred to give them a rating. This is confirmed by the fact that the non-native control group had, although not significantly, lower error rate in this condition than native speakers. Taken together, the non-native groups' results in this task indicate that they had indistinguishable performance.

5.2.1.3 Comparing Native Control to Non-Native Groups

The third procedure was to compare the native control group's ratings and mean percentage of errors in all conditions to those of the non-native groups. This was to find out if native speakers' superior knowledge helped them perform better than the non-native speakers. Mean ratings and percentage of errors of the three groups (i.e. native control, non-native control and non-native experimental) in the four conditions are compared in figure 5.5 and figure 5.6, respectively.

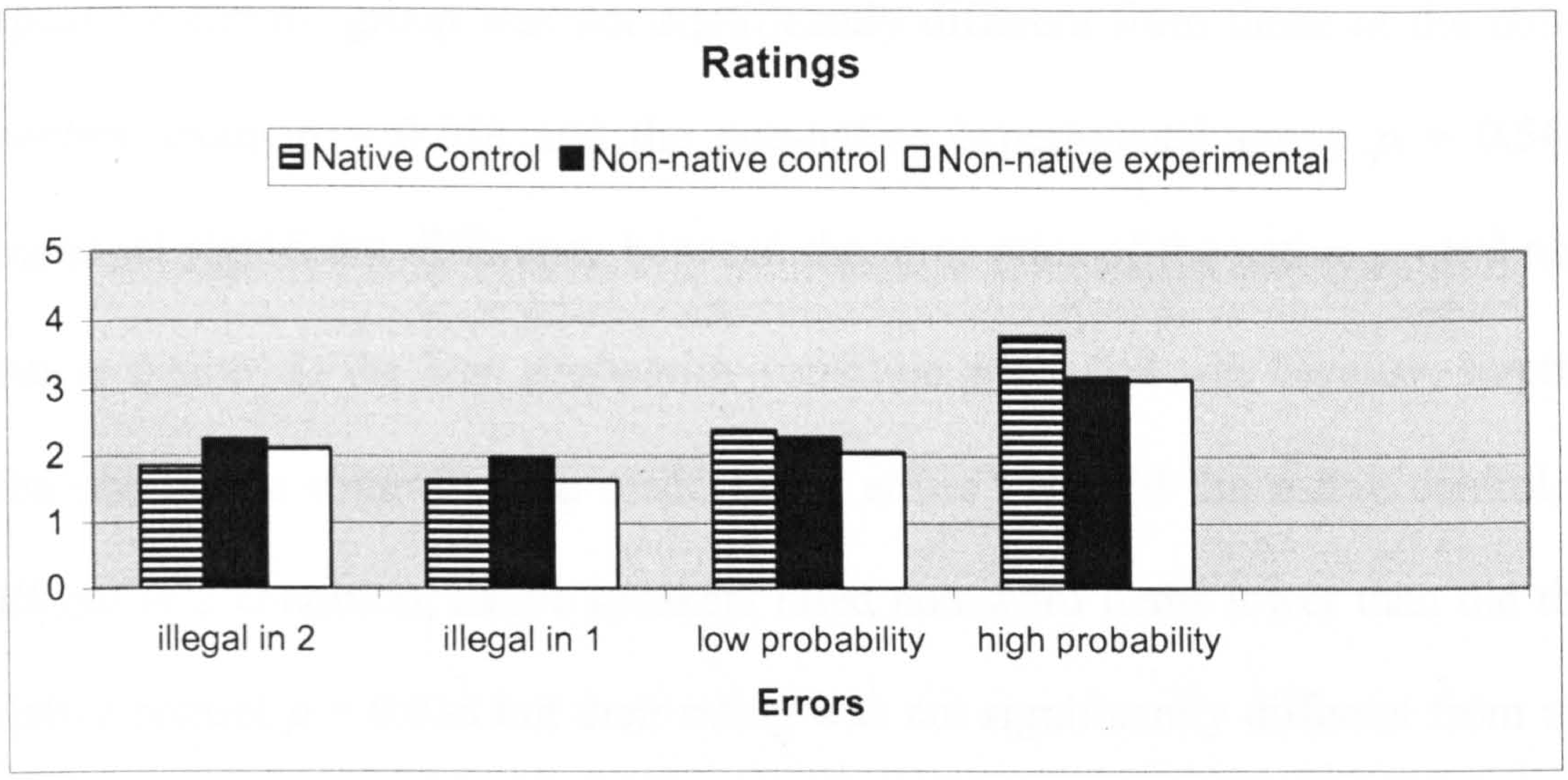

Figure 5.5. Mean ratings in the Non-word Rating Task of the three groups (native control, non-native control and non-native experimental) in the four phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability).

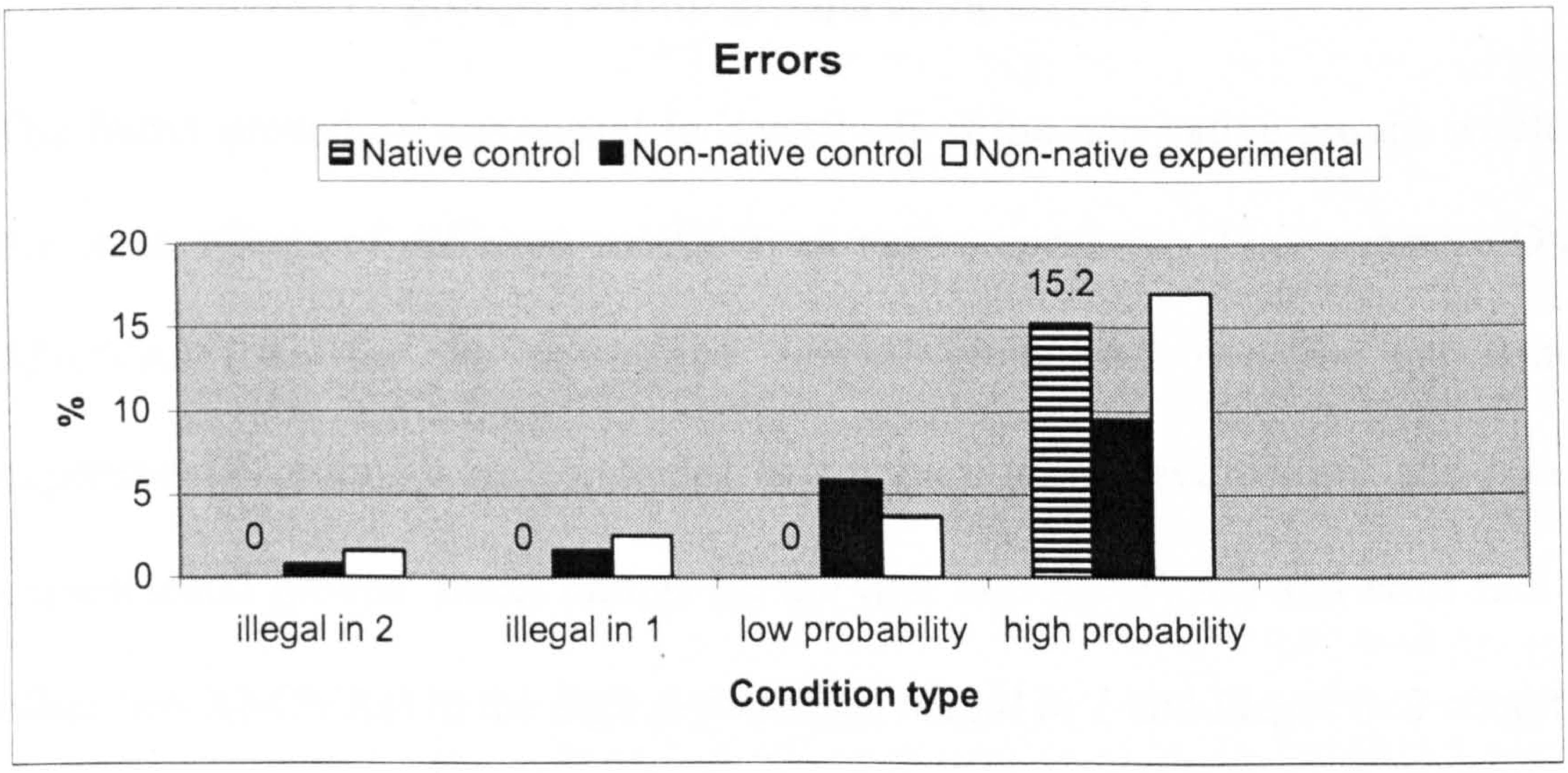

Figure 5.6. Mean percentage of errors in the Non-word Rating Task for the three groups (native control, non-native control and non-native experimental) in the four phonotactic conditions (illegal in 1, illegal in 2, low probability and high probability).

Two separate one-way ANOVAs were conducted to compare mean ratings in all conditions (in the first ANOVA) and error rate in the *high probability* condition (in the second ANOVA) of the native control group to those of the two non-native groups. Post hoc test results showed that the only condition where the native speakers' rating differed from those of the two non-native groups was in the *high probability* condition. In this condition, native English speakers rated non-words more English–like than did the non-native control group $p = 0.002$ and the non-native experimental group $p = 0.001$. However, in this condition, the error rate of the native English-speaking control group was not significantly different from those of the non-native control group $p = 0.058$ and the non-native experimental group $p = 0.546$. The marginal significant difference between the error rates of the native control and non-native control in the *high probability* condition $p = 0.058$ was because, surprisingly, the non-native control group made fewer errors than did the native control. In the *illegal in 2* condition, native speakers rated non-word items lower than did the non-native control $p = 0.038$ but their rating was not significantly different from the non-native experimental $p = 0.176$.

5.2.1.4 Non-native groups (within groups comparison)

The fourth procedure was meant to investigate if the non-native groups would show the same effects of different condition as native speakers. Four separate One-way ANOVAs (two for the non-native control group and two for the non-native experimental group) were conducted to compare non-native control and non-native experimental groups' mean ratings (in the first two ANOVAs) and error rate (in the other two ANOVAs) in the *high probability, illegal in 1* and *illegal in 2* conditions to those in the *low probability condition.* Post hoc test results showed that the two groups' ratings in the *high probability* condition were significantly higher than their ratings in the *low probability* condition ($p = 0.000$ for both groups). Also, the error rate was higher in the *high probability* condition than in the *low probability* condition only for the non-native experimental group $p = 0.000$ but not for the non-native control group $p = 0.110$.

Post hoc test results also showed that the non-native experimental group ratings and error rate in the *low probability* condition to those in the *illegal in 2* condition (Ratings, $p = 0.753$; Errors, $p = 0.501$). Their error rate was also statistically indistinguishable in the *low probability* condition from those in the *illegal in 1* condition $p= 0.686$. However, their ratings in the *illegal in 1* condition were significantly lower than their ratings in the *low probability* condition p= 0.026. Interestingly, their ratings, but not their error rate in the *illegal in 1* condition were also significantly lower than their ratings in the *illegal in 2* condition (Ratings, $p = 0.011$; Errors, $p = 0.788$).

On the other hand, post hoc test results showed that the non-native control group's ratings in the *low probability* condition were not significantly different from those in the *illegal in 2* condition $p= 0.835$. However, their error rate in the *illegal in 2* condition was significantly lower than those in the *low probability* condition $p = 0.033$. In the *illegal in 1* condition, both their ratings and error rate were lower than those in the *low probability* condition although the difference was marginally significant (Ratings, $p = 0.054$; Errors, $p = 0.074$). Their ratings, but not their error rate in the *illegal in 1* condition were also lower than their ratings in the *illegal in 2* condition although the difference was marginally significant (Ratings, $p = 0.085$; Errors, $p = 0.721$).

5.2.1.5 Discussion

Recall that the current Non-word Rating Task investigated two different research questions, namely subjects' sensitivity to the *legality* of English phonotactics and their sensitivity to the *probability* of English phonotactics. Therefore, the results relevant to each research question and related hypotheses will be discussed separately.

Sensitivity to phonotactic legality. Let us first recap the results of native speakers regarding their sensitivity to the legality of English phonotactics. In the Non-word

Rating Task, native speakers rated stimuli items in the *illegal in 2* condition (/tlauθ/) and *illegal in 1* condition (/dlɔɪθ/) less English-like than those in the *low probability* condition (/ðɔɪʒ/). However, there was no difference in error rates between these conditions. Therefore, this result partially confirmed Hypothesis 1, which states that "native speakers will rate the stimuli items in the *illegal in 2* and *illegal in 1* conditions lower and with fewer errors (i.e. classifying a non-word as a real word) than those in the *low probability* condition". Recall that the only difference between stimuli items in the *illegal in 2* and *illegal in 1* conditions and those in the *low probability* condition is that the former started with illegal clusters whereas the latter did not contain any illegal clusters but the stimuli were only made up of phonemes with low phonotactic probability. However, the phonotactic probability of the syllable starting at the second consonant of the non-word (e.g /laus/ in /dlaus/) in the *illegal in 2* and *illegal in 1* conditions were always low. It appears that what affected native speakers' judgement, making the difference in ratings, is merely the presence of initial illegal clusters.

These results show that native English speakers are sensitive to the phonotactic constraints of English and that this sensitivity can affect their subjective rating of non-words. These findings lend support to previous ones which showed that adult native speakers are affected by their knowledge of phonotactic constraints. As discussed in Chapter 3, phonotactic knowledge has long been shown to affect speech sound identification (e.g. Brown and Hildum 1956; Massaro and Cohen 1983; Pitt 1998). Phonotactic knowledge has also been shown to affect pronunciation, as mispronunciations usually follow the phonotactic constraints of the language in question (e.g. if /n/ is mispronounced as /ŋ/ it will be in a syllable coda) (Dell et al.

2000; Warker and Dell 2006). Finally, it has been shown that phonotactic knowledge can cause listeners to hear non-existent epenthetic vowels between phonemes when presented with consonant clusters that do not conform to the phonotactic constraints of their native language (Matthews and Brown 2004; Dupoux et al., 1999; Dupoux et al., 2001) as discussed in Chapter 3.

In addition, these results directly lend support to Coleman and Pierrehumbert's (1997) findings in that native speakers have reliable intuitions about phonotactic constraints in their own language. Their study was discussed in Chapter 3 but for reasons of convenience and because of their study yielded different results to those in the present Non-word Rating task, I repeat the presentation of their study here. Recall that Coleman and Pierrehumbert analysed native speakers' acceptability judgement of paired non-words starting with consonant clusters which either respected or violated phonotactic constraints (e.g. /glisləs/ and /mlisləs/). Subjects were asked to judge whether each item could or could not be a possible English word by pressing two response buttons. The number of *no* responses was counted and taken as a score of subjective degree of wordlikeness. Non-words starting with illegal clusters were significantly judged as less English-like than those starting with illegal ones. However, their analysis showed another interesting result. They found that some non-words starting with illegal consonant clusters (e.g. /mrupeiʃn/ were scored better than others starting with legal ones (e.g. /splɛtisak/).

Coleman and Pierrehumbert interpreted this finding by suggesting that acceptability judgement entails an evaluation of the whole composition of the non-word. High frequency or high phonotactic probability parts (e.g. tion in /mrupeiʃn/) will alleviate the ill-formedness of the illegal clusters. In brief, the presence of an illegal cluster is not enough to render a non-word unacceptable (ibid: 55).

Although a similar analysis of single items was not attempted in the current Non-word Rating Task,[54] general results are not compatible with Coleman and Pierrehumbert's finding. In the current Non-word Rating Task, native subjects rated stimuli items starting with illegal consonant clusters in the *illegal in 2* condition (/tlauθ/) and *illegal in 1* condition (/dlɔɪθ/) less English-like than those in the *low probability* condition (/ðɔɪʒ/). This is because, unlike the situation in Coleman and Pierrehumbert's task where frequency and probability of subsequent parts were not controlled, in the current task the phonotactic probability of the syllable starting at the second phoneme of the non-word (e.g. /lɔɪθ/) in /dlɔɪθ/) in the *illegal in 2* condition and *illegal in 1* condition was controlled and was always low. This suggests that when the rest of the item is controlled, native speakers, intuitions regarding the phonotactic constraints of their language can differentiate between low probability items and those which contain a subpart with zero probability (illegal clusters).

However, as the phonotactic probability of the items was controlled, the current task can not rule out Coleman and Pierrehumbert's proposal that judgements are based on evaluation of the whole composition of the non-word. Indeed, findings from the Lexical Decision Task are consistent with such a proposal. I will return to this issue when discussing results from the Lexical Decision Task.

---

[54] Such analysis is not relevant in the present study as the phonotactic probability of the syllable starting at the second phoneme of the non-word (e.g. /lɔɪθ/) in /dlɔɪθ/) in the *illegal in 2* condition and *illegal in 1* condition was controlled.

Let us now consider non-native subjects' results. Do they show sensitivity to phonotactic legality of English and Arabic as measured by the Non-word Rating Task when listening to English? Null Hypothesis 3, which states that "non-native speakers will show no difference between their ratings or error rates in the *illegal in 2* condition, *illegal in 1* condition and the *low probability* condition", was partially rejected. First, both non-native groups judged stimuli items in the *illegal in 2* (/tlauθ/) as as English-like as those in the *low probability* condition (/ðɔɪʒ/) condition. Only the non-native control group, however, made fewer errors when rating stimuli items in the *illegal in 2* condition than those in the *low probability* condition. On the other hand, both non-native groups judged stimuli items in the *illegal in 1* condition (/dlɔɪθ/) to be less English-like than those in the *low probability* condition (see Figure 5.3). In addition, the non-native control group made fewer errors when rating stimuli items in the *illegal in 1* condition than those in the *low probability* condition (see Figure 5.4).

Ratings of both groups in the *illegal in 1* condition compared to their ratings in the *low probability* condition suggest that they are sensitive to the phonotactic constraints of English. They consistently judged stimuli items in the *illegal in 1* condition to be less English-like than those in the *low probability* condition. This result lends support to previous findings (e.g. Altenberg and Cairns 1983; Altenberg 2005b) that bilinguals and L2 learners' subjective ratings show good knowledge of the legality of consonant clusters in English. However, surprisingly, our non-native subjects did not judge stimuli items in the *illegal in 2* condition to be less English-like than those in the *low probability* condition. Their sensitivity to the phonotactic constraints of English and Arabic in the *illegal in 2* condition showed only as lower error rates in this condition in the non-native control group's ratings compared to those in the *low probability* condition.

The most interesting and unpredicted result, however, was that both non-native groups judged stimuli items in the *illegal in 1* condition to be less English-like than those in the *illegal in 2* condition. Recall that in addition to being illegal in Arabic, stimuli items in the *illegal in 2* condition are also illegal in English, as those in the *illegal in 1* condition. Why should this happen? Non-native speakers were actually predicted to judge stimuli items in the *illegal in 2* condition to be less English-like or at least as English-like as those in the *illegal in 1* condition. One possible explanation is that because non-native subjects were told that they would listen to English words and non-words composed of English phonemes, and given the time this task allowed them, they managed to judge these items based on their knowledge of English only, in other words switching to a monolingual English mode.

Grosjean (2001) mentions a number of factors that could affect the language mode. These include stimuli, situation and participants' knowledge that they are participating because they are bilingual, which may in fact encourage them to operate in a bilingual mode and consequently influence their responses. Actually, the procedure followed in our study encouraged subjects to be in a monolingual English mode. As discussed in Chapter 4, our subjects were EFL learners who likely did not know that there was a particular interest in them as speakers of Arabic. In addition, recall that the stimuli in the *low probability* and *high probability* conditions were mostly constructed of non-L1 phonemes, another factor which may have encouraged non-native subjects to judge stimuli items based solely on their knowledge of English.

However, a closer examination of our non-native subjects' behaviour indicates that they were not merely operating in a monolingual English mode. Rather, it seems that they were deliberately trying to do so by avoiding basing their judgement on their knowledge of Arabic phonotactic constraints and the Non-word Rating Task gave them enough time to do that. This is because they judged stimulus items starting with illegal consonant clusters in English only as less English-like than those starting with illegal consonants clusters in both English and Arabic.

Similar findings were found by Altenberg and Cairns (1983) in a visual presentation of the same task. When they asked English monolinguals and English-German bilinguals to rate how English-like non-words starting with illegal consonant clusters in German only (e.g. smatt), both groups' ratings were only influenced by phonotactic constraints of English. However when the same stimulus items where presented in a visual Lexical Decision Task, processing times of the bilinguals were affected by the phonotactic constraints of German. Similarly, Altenberg (2005b) found in a rating task that Spanish ESL learners judged non-words starting with illegal clusters in Spanish only more English-like than those starting with illegal clusters in English and Spanish, suggesting that their judgements were not affected by the status of these clusters in the L1.

Another plausible explanation concerns the effect that phonotactic constraints can have on speech sound identification. Recall that researchers such as Brown and Hildum (1956); Massaro and Cohen (1983) and Pitt (1998) have shown that native speakers' labelling of an ambiguous segment, usually in less than perfect conditions, is influenced by their native language phonotactic permissibility of the consonant cluster containing that ambiguous sound. When subjects were asked to label an ambiguous segment between [l] and [r], for example, the ambiguous segment was more often labelled as [r] in [t?i], while the same segment was more often labelled as [l] in [s?i] (Massaro and Cohen 1983).

It seems that being a non-native listener resembles listening to one's native language in less than perfect conditions. In other words, it is possible that, influenced by their L1 phonotactic constraints, our non-native listeners misperceived the non-L1 consonant clusters in the *illegal in 2* condition (e.g. /sr/ in /srud/). Non-native listeners may have misperceived the illegal cluster /sr/ as the legal one [ʃr/] or [fr] rendering it legal in English as well. On the other hand, their knowledge of English phonotactic constraints may not be subtle enough to cause such misidentification in the *illegal in 1* condition. This could explain why non-native listeners' ratings of stimuli items in the *illegal in 2* condition were statistically indistinguishable from those in *low probability* conditions and higher than those in the *illegal in 1* condition.

Sensitivity to phonotactic probability. Let us now recap the results of native speakers regarding their sensitivity to the probability of English phonotactics. In the Non-word Rating Task, native speakers judged stimulus items in the *high probability* condition as more English-like than those in the *low probability* condition. They even made some errors when rating stimulus items in the *high probability* condition (i.e. they wrongly classified some high probability non-words as real words). This result confirms Hypothesis 2, which states that "Native speakers will rate the stimuli items in the *low probability* condition lower and with fewer errors than those in the *high* *probability* condition". Additionally, this result lends support to previous findings that English native speakers are sensitive to the phonotactic probability in their L1 and that this sensitivity could show in their subjective ratings of non-words (Bailey and Hahn 2001; Frisch et al. 2000; Vitevitch et al. 1997), as discussed in Chapter 3. This result is consistent with the claim that phonotactic probability information stored in memory can be accessed by native speakers and used to give reliable word-likeness judgments of non-word stimuli (Vitevitch et al. 1997).

However, results from native speakers do not provide information regarding the source of phonotactic probability knowledge, that is "whether listeners have access to independent information in memory regarding phonetic segments or sequences, or whether all phonotactic effects emanate from individual representation of lexical form" (Vitevitch et al. 1997: 48). In addition, as our stimuli was either high probability/high density or low probability/ low density, it can not be ensured that the effect obtained in the judgments of native speakers in the current task is the result of phonotactic probability, although previous studies (e.g. Bailey and Hahn 2001; Vitevitch and Luce 1998, 1999, 2005) have shown that phonotactic probability has a significant effect that is beyond and independent of neighbourhood density.

As I argued in Chapter 3, the difficulty of dissociating the effect of phonotactic probability and neighbourhood density in native speakers lies in large part in the fact that adult native speakers have completely constructed lexicons. Therefore, our native speakers could have been affected by neighbourhood density of the stimuli as well, judging words in the high probability/high density condition more English-like because they find them similar to more English words in their lexicon. This is where results from our non-native speakers are vital. EFL learners do not have at all as complete an English lexicon as English native speakers. Thus, the lexicon, and more specifically neighbourhood density effects, are not expected to play the same role as in native speakers because of the different distribution of words in the native and non-native lexicons. The scale of density effects here depends on the number of neighbours acquired by the learner.[55]

As mentioned in Chapter 3, Common but low density words such as *face, dog, home, book, wife, room* appear earlier in teaching materials than uncommon but high density ones (e.g. *muss, ram, soar* etc..). The main point here is that a non-native's lexicon is not as informative as a native speaker's one as far as its effect on sensitivity is concerned. Therefore, if subjects in the current study (low-intermediate EFL learners) show sensitivity to stimuli that is high probability/ high density, this would provide additional evidence that phonotactic probability effects are unique and are not only subsumed by neighbourhood density effects. For the same argument regarding the incompatibility of native and nonnative lexicons discussed above, if our EFL subjects show sensitivity to the phonotactic probability in the Non-word Rating Task, this would be suggestive that this sensitivity stems from independent information in memory regarding phonotactic probability rather than from individual representation of lexical form (Vitevitch et al. 1997).

Hypothesis 4, which states that "non-native speakers will show no difference between their ratings or error rates in the *high probability* condition and the *low probability*

---

[55] The number of neighbours of a certain stimulus item which our low intermediate non-native speakers have acquired is likely to be much fewer than those acquired by native speakers and hence a high level of sensitivity of our non-native speakers to high density/high probability items can not be attributed to neighbourhood density effects alone.

condition" was rejected as our non-native speakers seem to be sensitive to the phonotactic probability of English. Subjects in both non-native groups judged stimuli items in the *high probability* condition more English-like than those in the *low probability* condition. In addition, the non-native experimental group made more errors when judging stimuli items in the *high probability* condition than in the *low probability* condition (i.e. they wrongly classified some *high probability* non-words as real words). These results provide further evidence that phonotactic probability has an effect that is beyond and independent of neighbourhood density. It also seems that phonotactic probability effects are the result of abstract phonological information in memory that is independent of lexical representations of sound patterns. This latter proposal is consistent with the Shortlist Model of Word Recognition (Norris 1994). Unlike TRACE (McClelland and Elman 1986), which suggests that phonotactic effects are top-down lexical effects, Shortlist suggests that phonotactic information is independently represented knowledge.

Other indirect support for this proposal comes from subjects' (both native and non-native) sensitivity to the phonotactic legality of English. Here sensitivity can not be the result of the presence of these constraints in lexical items; rather it is the result of the absence of these sequences (illegal clusters) in lexical items, which is suggestive of the availability of an abstract knowledge of phonotactic constraints. Findings from infants are also supportive of the proposal that phonotactic probability effects are the result of abstract phonological information in memory that is independent of lexical representations of sound patterns. As Jusczyk (1997) noted, infants seem to develop sensitivity to the sound patterns of their native language before "significant vocabulary development" because comprehending first words is evidenced between

8-10 months of age (ibid: 89). This is evidence which favours the proposal that this sensitivity does not result from "extracting this information from large number of known words" since babies do not recognise any when they exhibit this sensitivity (ibid).

5.2.2 Results of the Lexical Decision Task

In this task two dependent variables were analysed, namely reaction times and error rate. An error was counted when a 'no' response was made when a non-word was presented. Practice items and real-word fillers were not included in the analysis. The error rate was not analysed as a dependent variable in previous research, which was mostly with native speakers. However, given the level of our non-native subjects, and as is the case with the Non-word Rating Task, different conditions in this task were predicted to show effects in errors, hence the analysis of error rate in this task. In addition, because of the time constraint which this task entailed, it was predicted that effects of different conditions would show in error rates in this task more than they did in the Non-word Rating Task.

Recall that the current task investigated two different research questions, namely subjects' sensitivity to the *legality* of English phonotactics and their sensitivity to the *probability* of English phonotactics. As with the Non-word Rating Task, previous research using the Lexical Decision Task has tried to answer one question or the other, but not both. Therefore, in the current task four different conditions, instead of two, were compared: *illegal in one* (English only), *illegal in both* (English and Arabic), *low phonotactic probability* and *high phonotactic probability*.

The best way to answer our two research question is to compare RTs and error rates in all conditions to the *low phonotactic probability* condition. To answer the first research question, RTs and error rates in the *illegal in one* condition and *illegal in two* condition will be compared to those in the *low phonotactic probability* condition. If subjects are sensitive to the legality of English phonotactics they are predicted to reject non-words in the *illegal in one* and *illegal in two* conditions faster and with fewer errors than those in the *low phonotactic probability* condition. To answer the second research question, RTs and error rate in the *high phonotactic probability* condition will also be compared to those in the *low phonotactic probability* condition. If subjects are sensitive to the phonotactic probability of English, they are predicted to reject non-words in the *high phonotactic probability* condition more slowly and with more errors than those in the *low phonotactic probability* condition. In the following discussion, effects are considered statistically significant at $p \leq 0.05$ and as marginally significant where $0.05 < p \leq 0.1$. (a) Mean RTs in milliseconds (ms) is measured from item onset and mean percentage of error rate for the three subjects groups (native control, non-native control and non-native experimental) in the four conditions (high probability, low probability, illegal in 1 and illegal in 2) are shown here table 5.1.

Table 5.1 Mean RTs in ms and mean percentage of errors for the 3 groups in phonotactic conditions.

Condition type	Illegal in 2	Illegal in 1	Low probability	High probability
(a) RTs in ms				
Native speakers	1205	1060	1271	1594
Non-native control	1807	1561	1743	2075
Non-native experimental	1823	1546	1752	2066
(b) % of Errors				
Native speakers	6.9	2.7	5.5	24.3
Non-native control	21.6	16.6	16.6	43.7
Non-native experimental	27.5	14.1	19.1	39.5

5.2.2.1 Native control

The first procedure was to find out if our native control group showed effects of different conditions. Their mean RTs in milliseconds measured from item onset and mean percentage of error rate for each condition are compared in Figure 5.7 and Figure 5.8 respectively.

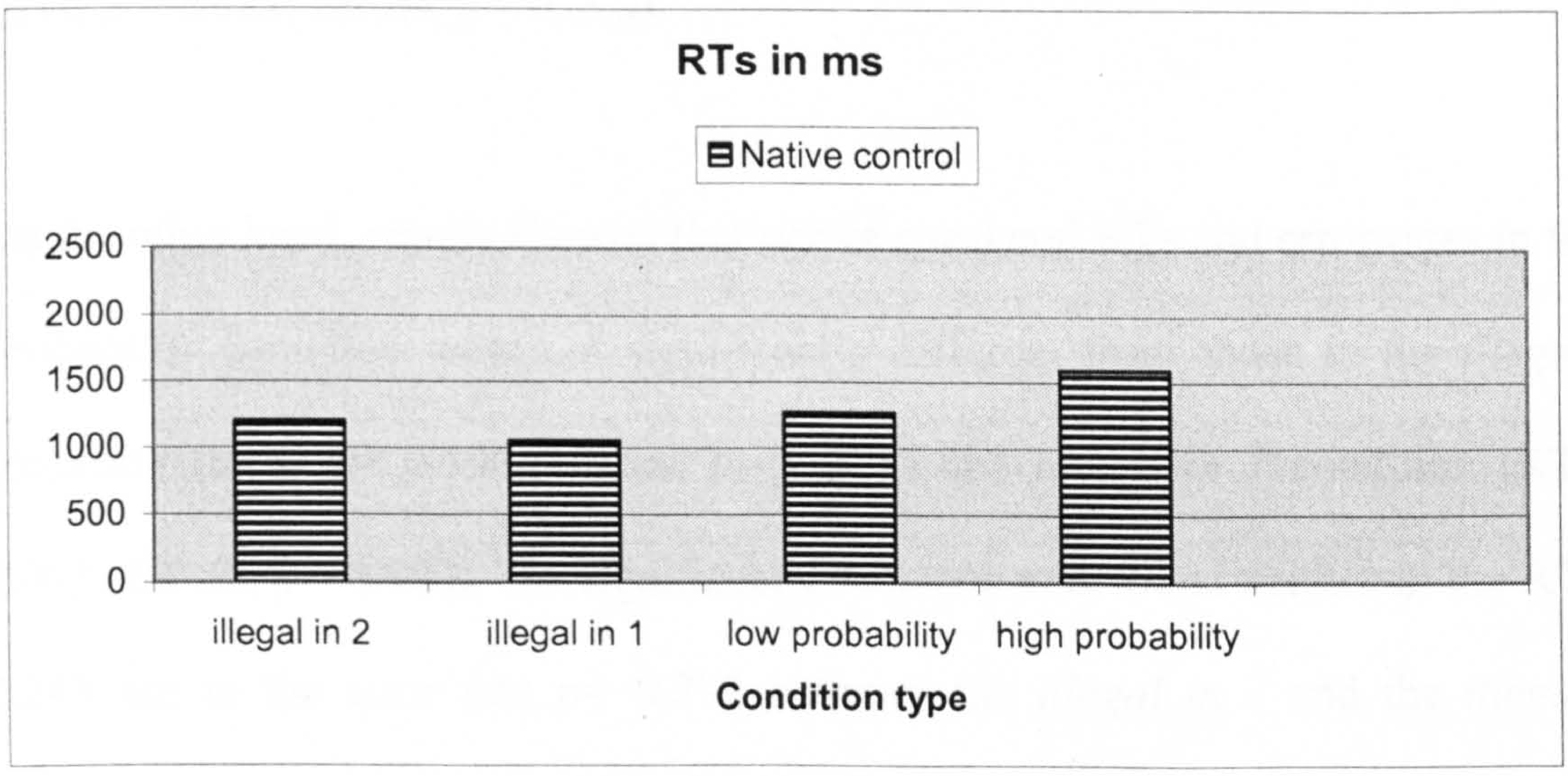

Figure 5.7 Native speakers' mean RTs in ms measured from item onset in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability).

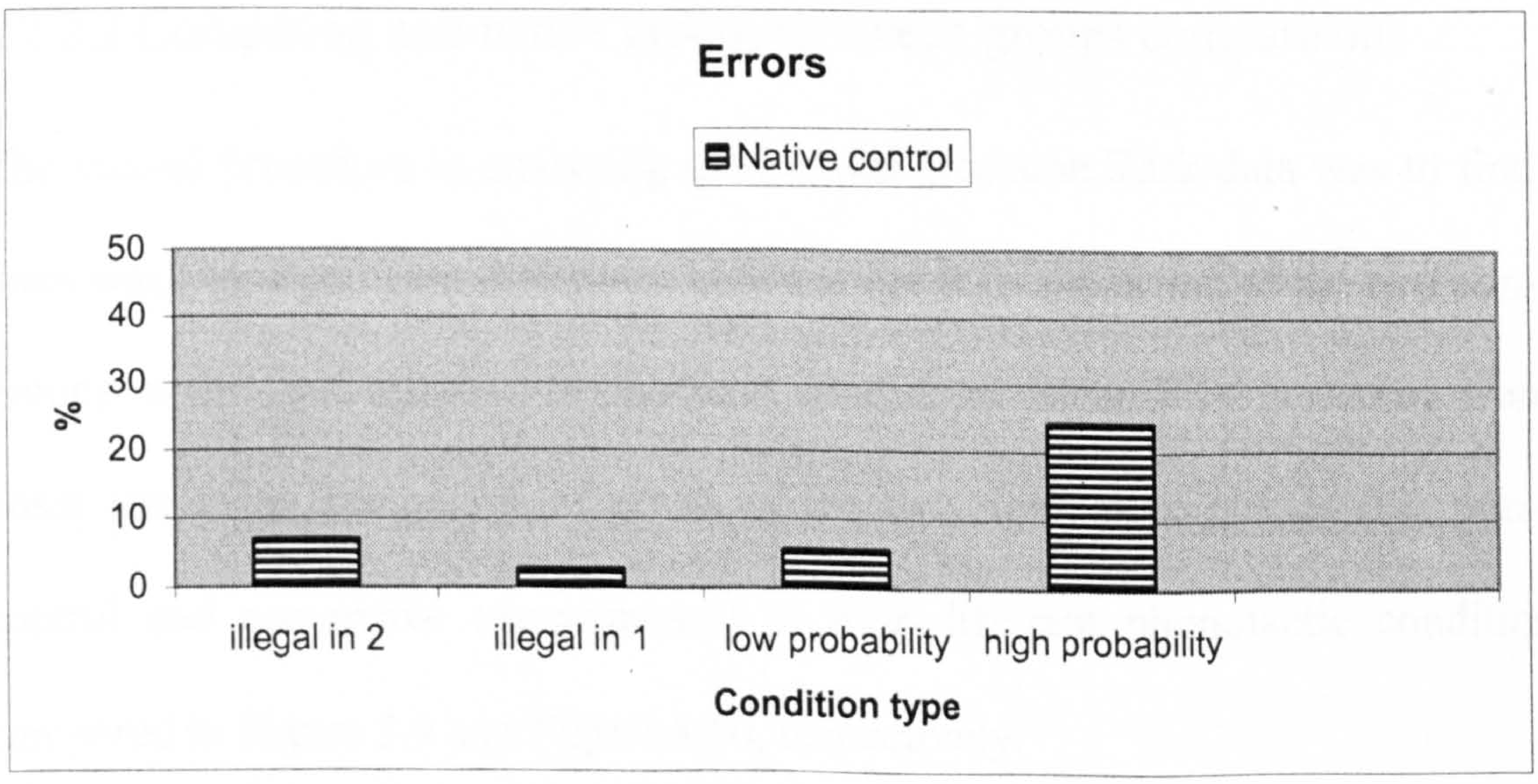

Figure 5.8 Native speakers' mean percentage of errors in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability).

As Figure 5.8 shows, unlike the Non-word Rating Task where native speakers committed errors only in the *high probability* condition, in the Lexical Decision Task native speakers committed errors in all conditions. Two separate One-way ANOVAs were conducted to compare native speakers' mean RTs (in the first ANOVA) and error rate (in the second) in the *high probability, illegal in 1* and *illegal in 2* conditions to those in the *low probability condition.* Post hoc test results showed that native speakers' RTs and error rate in the *high probability* condition were significantly higher than their RTs and error rates in the *low probability* condition (RTs, $p = 0.012$; Errors, $p = 0.000$).

On the other hand, results showed that native speakers' RTs and error rates in the *low probability* condition were not significantly different from those in the *illegal in 2* condition (RTs, $p = 0.596$; Errors, $p = 0.724$) and *illegal in 1* condition (RTs, $p = 0.093$; Errors, $p = 0.480$). No significant difference was found neither in the RTs $p = 0.243$ nor in the error rate $p = 0.292$ between the *illegal in 2* and the *illegal in 1* conditions.

5.2.2.2 Comparing non-native groups (between groups comparison)

The second procedure in analysing the Lexical Decision Task data was to find out if there was any significant difference between the RTs and errors of the two non-native groups in the four different phonotactic conditions. Mean RTs measured from item onset and mean percentage of errors of the two non-native groups (i.e. non-native control and non-native experimental) in four different phonotactic conditions are compared in Figure 5.9 and Figure 5.10, respectively.

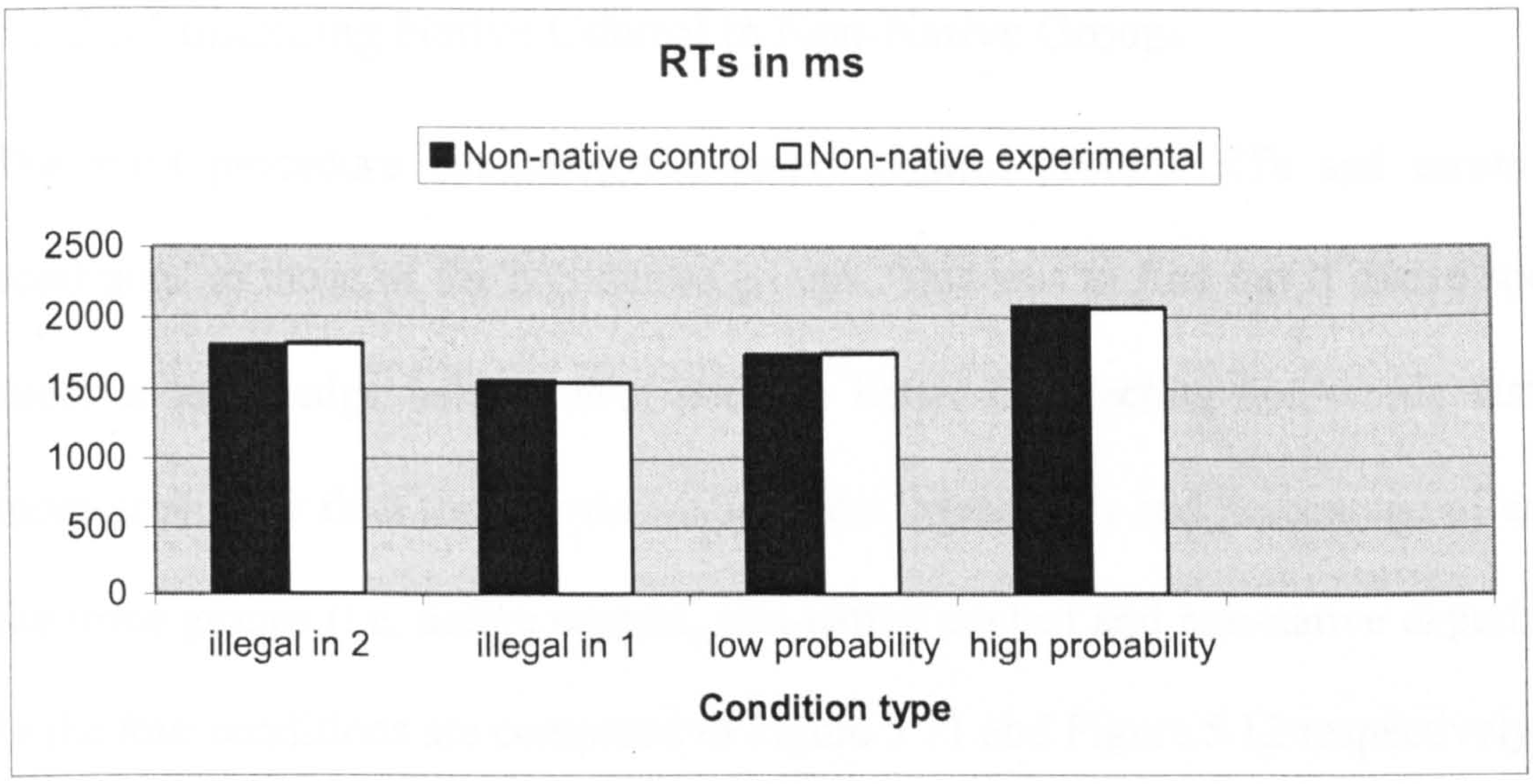

Figure 5.9 Non-native control and non-native experimental groups mean RTs in ms measured from item onset in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability).

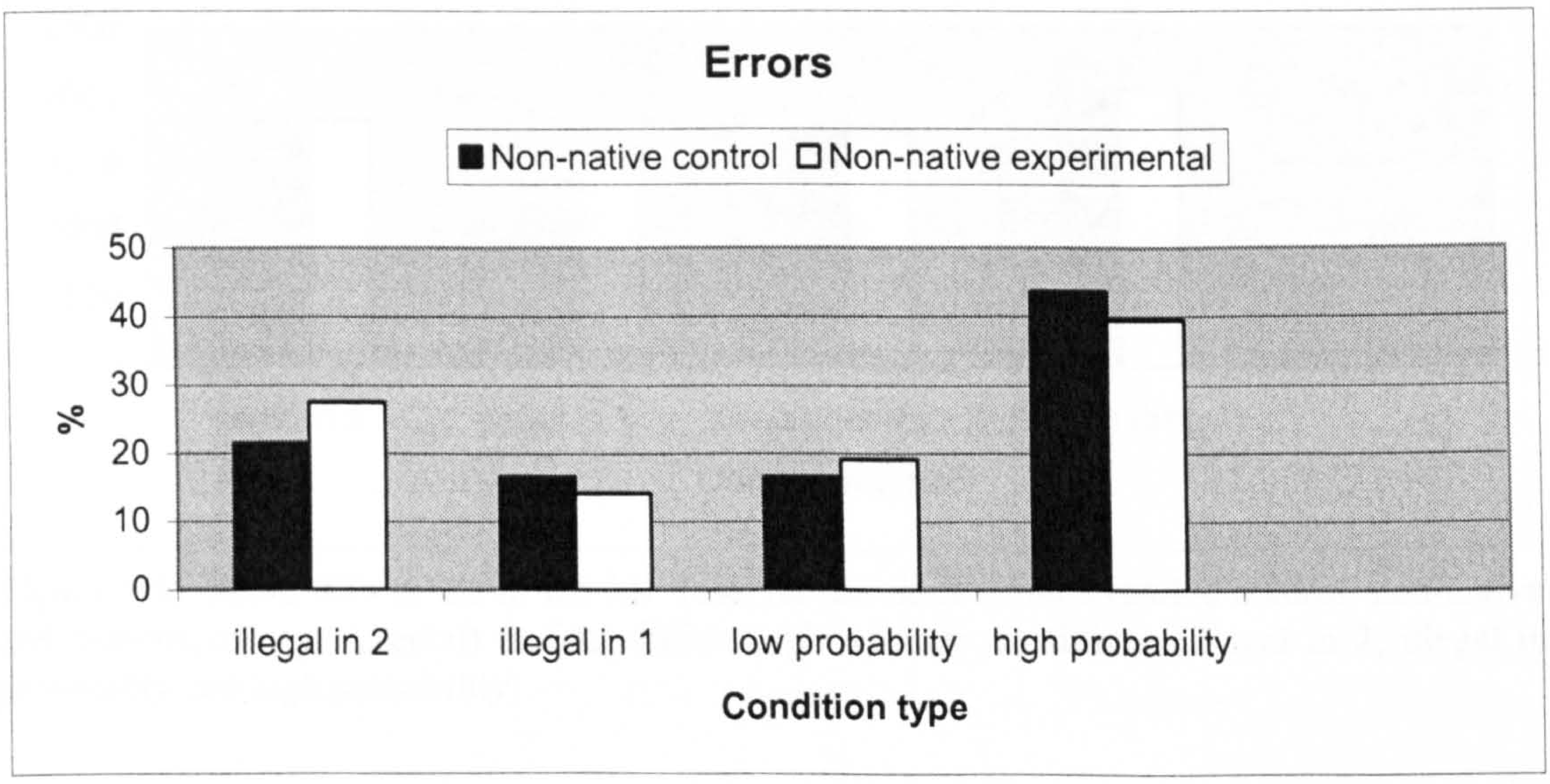

Figure 5.10 Non-native control and non-native experimental groups' mean percentage of errors in the Lexical Decision Task in four different phonotactic conditions (illegal in 2, illegal in , low probability and high probability).

Two separate One-way ANOVAs were conducted to compare mean RTs (in the first ANOVA) and error rate (in the second) of the two non-native groups in all four conditions. Post hoc test results showed that there was no significant difference between the two groups' RTs and error rates in all four conditions; *illegal in 2* condition (RTs, p= 0.879; Errors, p= 0.272), *illegal in 1* (RTs, p= 0.903; Errors, p= 0.637), *low probability* (RTs, p= 0.942; Errors, p= 0.637) and *high probability* (RTs, p= 0.944; Errors, p= 0.432).

5.2.2.3 Comparing Native Control to Non-Native Groups

The third procedure compared the native control group's RTs and errors in all conditions to those of the non-native groups. This was to find out if native speakers' superior knowledge helped them perform better in rejecting non-words faster and more accurately than the non-native speakers. Mean RTs and percentage of errors of the three groups (i.e. native control, non-native control and non-native experimental) in the four conditions are compared in Figure 5.11 and Figure 5.12 respectively.

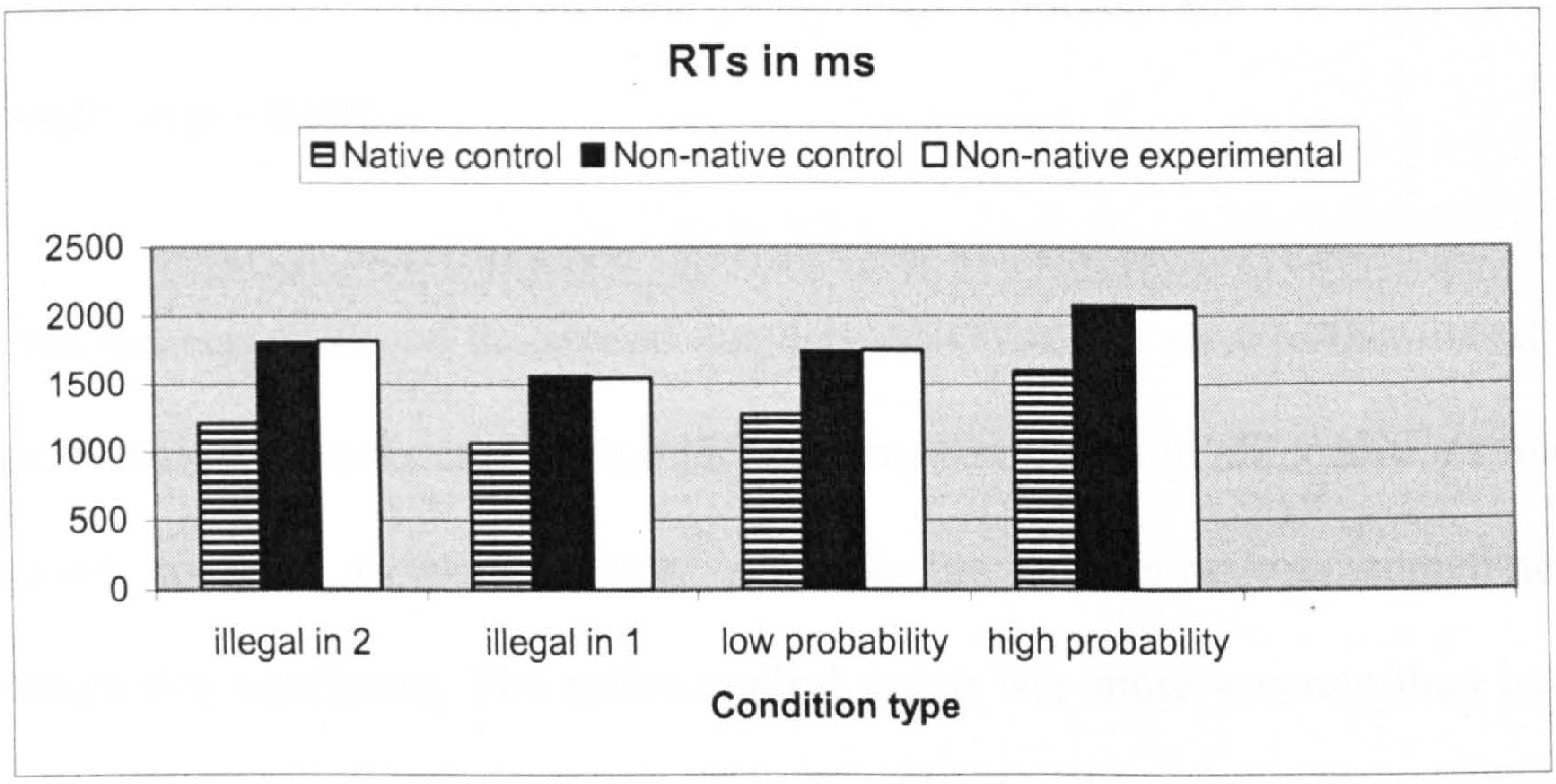

Figure 5.11 Mean RTs in ms in the LD Task for the three groups (native control, non-native control and non-native experimental) in four different phonotactic conditions (illegal in 2, illegal in 1, low probability and high probability).

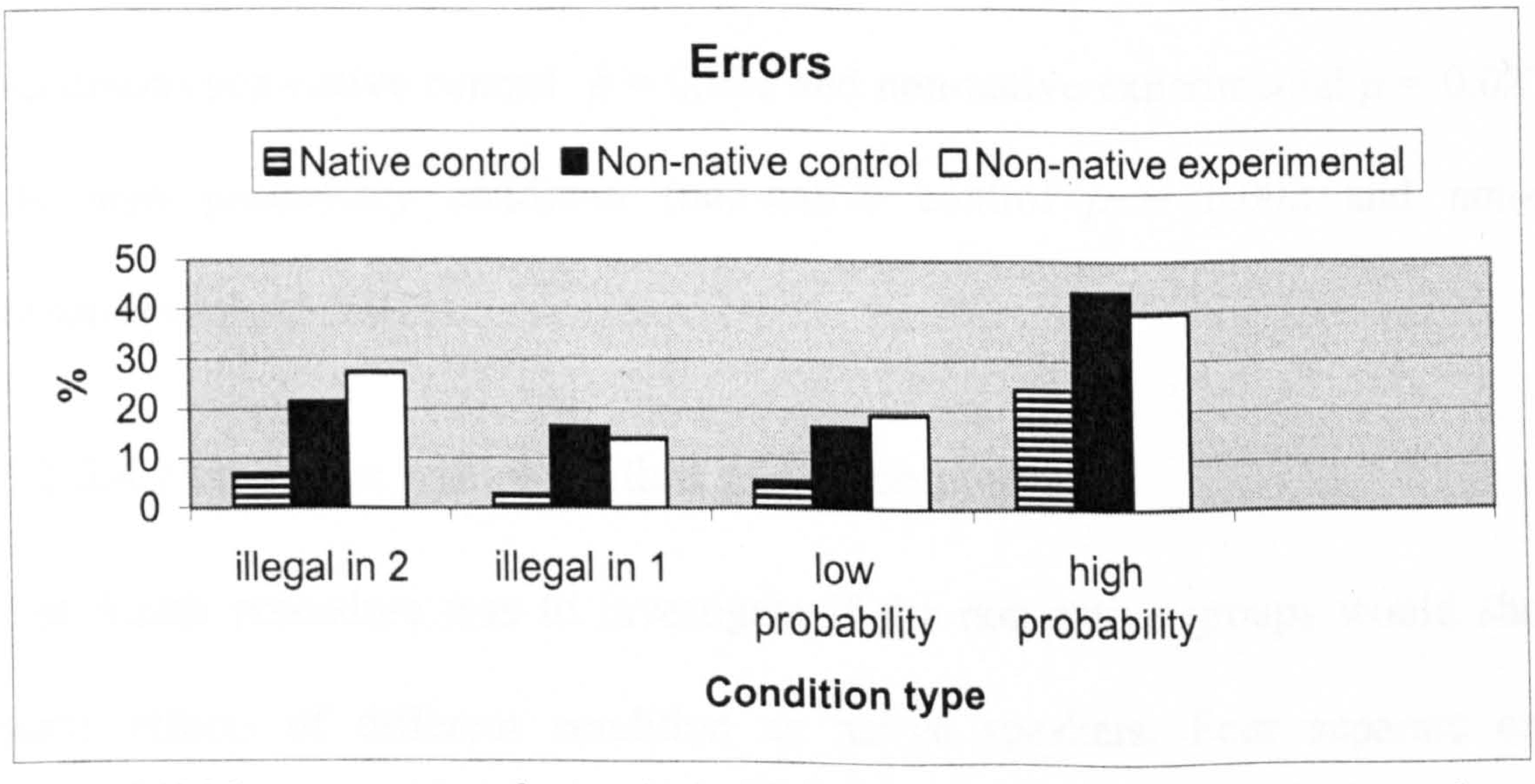

Figure 5.12 Mean percentage of errors in the LD Task for the three groups (native control, non-native control and non-native experimental) in the four different phonotactic conditions).

Two separate one-way ANOVAs were conducted to compare mean RTs (in the first ANOVA) and error rate in (in the second ANOVA) in all conditions of the native control group to those of the two non-native groups. Post hoc test results of the first one-way ANOVA showed that native speakers were faster in rejecting non-word items in all conditions than non-native speakers in both groups. The native control group was faster in rejecting non-words than both groups in the *illegal in 2* condition $p = 0.000$, the *illegal in 1* condition (non-native control $p = 0.000$ and non-native experimental $p = 0.001$), *the low probability* condition and *the high probability* condition $p = 0.001$.

Post hoc test results of the second one-way ANOVA also showed that overall native speakers were more accurate in rejecting non-word items in all conditions than non-native speakers in both groups, although the difference was sometimes only marginally significant. The native control group was more accurate than both non-native groups in rejecting non-words in the *illegal in 2* condition (non-native control $p = 0.017$ and non-native experimental $p = 0.001$), the *illegal in 1* condition (non-native control $p = 0.024$ and non-native experimental $p = 0.064$), *the low probability* condition (non-native control $p = 0.071$ and non-native experimental $p = 0.027$) and *the high probability* condition (non-native control $p = 0.002$ and non-native experimental $p = 0.013$).

5.2.2.4 Non-native groups (within groups comparison)

The fourth procedure was to investigate if the non-native groups would show the same effects of different condition as native speakers. Four separate one-way ANOVAs (two for the non-native control group and other two for the non-native experimental group) were conducted to compare non-native control and non-native experimental groups' mean RTs (in the first two ANOVAs) and error rate (in the other two ANOVAs) in the *high probability, illegal in 1* and *illegal in 2* conditions to those in the *low probability condition.* Post hoc test results showed that the two groups' RTs in the *high probability* condition were significantly higher than their RTs in the *low probability* condition (non-native control p = 0.012 and non-native experimental p = 0.018). The error rate was also higher in the *high probability* condition than in the *low probability* condition for both groups (non-native control p = 0.000 and non-native experimental p = 0.001).

On the other hand, post hoc test results showed that both non-native groups' RTs and error rates in the *low probability* condition were not significantly different from those in the *illegal in 2* condition (Non-native control, RTs, p= 0.558; Errors, p= 0.100; non-native experimental, RTs, p= 0.585; Errors, p= 0.167). Similarly, both non-native groups' RTs and error rates in the *low probability* condition were not significantly different from those in the *illegal in 1* condition (Non-native control, RTs, p= 0.161; Errors, p= 0.651; non-native experimental, RTs, p= 0.115; Errors, p= 0.405). Interestingly, both groups' RTs and error rates in the *illegal in 1* condition were lower than their RTs and error rates in the *illegal in 2* condition (Non-native control, RTs, p= 0.049; Errors, p= 0.037; non-native experimental, RTs, p= 0.035; Errors, p= 0.029).

5.2.2.5 Discussion

Recall that the current Lexical Decision Task investigated two different research questions, namely subjects' sensitivity to the *legality* of English phonotactics and their sensitivity to the *probability* of English phonotactics. The results relevant to each research question will be discussed separately. These are the same research questions the Non-word Rating Task investigated. The main aim of using the Lexical Decision Task, however, was to find out if sensitivity shown in a task where metalinguistic knowledge could be used such as the Non-word Rating and could also be shown in an online task where subjects are denied the time element. In other words, the aim of the Lexical Decision Task was to find out as Vitevitch et al. (1997: 55) put it "if measures of processing time coincide with subjective ratings of phonological goodness" such that non-words judged as most English-like in the Non-word Rating Task will be processed and rejected most slowly and vice versa in the Lexical Decision Task.

Sensitivity to phonotactic legality. Let us first recap the results of native speakers regarding their sensitivity to the legality of English phonotactics. In the Lexical Decision Task native speakers rejected non-word stimuli items in the *low probability* condition as fast and with a statistically indistinguishable number of errors as they did with those in the *illegal in 2* condition and *illegal in 1* condition. Therefore, our Hypothesis 5, which states that "Native speakers will reject stimuli items in the *illegal in 2* and *illegal in 1* conditions faster and with fewer errors than those in the *low probability* condition", is disconfirmed. Native speakers were also faster and more accurate than non-native speakers in both groups in deciding that non-words are non-words in the *illegal in 2* and *illegal in 1* conditions. This result partially confirms Hypothesis 7, which states that "Native speakers will be faster in rejecting stimuli items and with fewer errors in all conditions than non-native speakers".

Note that the only difference between stimuli items in the *illegal in 2* and *illegal in 1* conditions on the one hand and those in the *low probability* condition on the other is that the former started with illegal clusters. However, the phonotactic probability of the syllable starting at the second consonant of the non-word (e.g /laus/ in /dlaus/) in the *illegal in 2* and *illegal in 1* conditions was always low. In the Non-word Rating Task native speakers' judgements were affected by this difference (i.e. starting with illegal clusters) and therefore they rated the stimuli items in the *illegal in 2* and *illegal in 1* conditions lower than those in the *low probability* condition. But why has this difference not shown an effect in the lexical Decision Task?

Recall from our discussion in Chapter 3 and in Section 5.2.1.5 of this chapter that Coleman and Pierrehumbert (1997) found that some non-words starting with illegal consonant clusters (e.g. /mrupeiʃn/ were rated better than others starting with legal ones (e.g. /splɛtisak/). They interpreted this finding by suggesting that acceptability judgement entails an evaluation of the whole composition of the non-word. Because the part following the illegal cluster in our stimuli was controlled, our findings from the Non-word Rating Task did not support such a proposal as our native speakers consistently judged stimuli items starting with illegal clusters as less English-like than those with low phonotactic probability.

However, findings from the current Lexical Decision Task lend support to a proposal that timed lexical decisions, but not the subjective ratings Coleman and Pierrehumbert propose, are based on an evaluation of the whole composition of the non-word. A closer examination of the procedure which Coleman and Pierrehumbert followed in their acceptability judgement task shows us why our findings from the Lexical Decision Task supported their proposal but our findings from the Non-word Rating Task did not. Coleman and Pierrehumbert did not arrive at subjective ratings by providing subjects with a scale and asking them to rate stimuli items on that scale as in the current Non-word Rating Task. Rather, they aurally presented their subjects with the stimuli items and asked them to judge whether each word could or could not be a possible English word by pressing one of two response buttons. Their score of subjective degree of well-formedness was based on the total number of responses against the well-formedness of each item. An item given 12 responses against (the highest score) was considered completely unacceptable. This kind of procedure made their task more like a Lexical Decision Task in which, as our Lexical Decision Task has shown, subjects do not reject non-words based only on the initial illegal cluster but listen to whole non-words to evaluate the entire compositions of the item. Consequently RTs did not show an effect of the presence of illegal clusters.

On the other hand, it seems that the same thing happened in the current Non-word Rating Task (i.e. listening to the whole item), but because of the time subjects are allowed they are able to process the item again by going back to its beginning and ultimately base their judgements on the presence of that illegal cluster. This is a possible reason why the presence of an illegal consonant cluster has shown an effect only in the Non-word Rating Task but not in the Lexical Decision Task. Therefore, it seems that on-line tasks such as The Lexical Decision Task are not appropriate for investigating speakers' sensitivity to the phonotactic legality of their language as represented in stimuli items starting with illegal clusters. This finding provides evidence against the proposal that words are recognised at their uniqueness points (e.g. Marslen-Wilson 1973; Marslen-Wilson and Welsh 1978). As discussed in Chapter 3, an ideal recognition point predicts that words can become unique and consequently be recognized before their acoustic offsets. The implausibility of an ideal recognition point was also confirmed by empirical evidence from studies discussed in Chapter 3. Using a gating task, Grosjean (1985) showed that many short words could only be recognized some time after their offsets. Similarly, Goodman and Huttenlocher (1988) used a Lexical Decision Task and varied when during the sound pattern (i.e. non-words) the deviation point occurred. They found that reaction times measured from the deviation point were longer when the deviation point occurred earlier in the non-word suggesting that information after the isolation point was needed for identification.

Non-native speakers' results also suggest that in general the Lexical Decision Task failed to show that non-native speakers are sensitive to the phonotactic legality of either English or Arabic. Null Hypothesis 8, which states that "Non-native speakers will show no difference in RTs or error rates in rejecting stimuli items in the *illegal in 2* condition, *illegal in 1* condition and *low probability* condition" was partially accepted. Our Non-native speakers in both groups rejected non-word stimuli items in the *low probability* condition as fast and with statistically indistinguishable number of errors as they did with those in the *illegal in 2* condition and *illegal in 1* condition. Recall that in the Non-word Rating Task, non-native speakers have shown that they are sensitive to the phonotactic legality of English and to a lesser degree to that of Arabic. On the one hand, the non-native control group made fewer errors when rating stimuli items in the *illegal in 2* condition than those in the *low probability* condition. On the other hand, both non-native groups judged stimuli items in the *illegal in 1* condition (/dlɔɪθ/) to be less English-like than those in the *low probability* condition. In addition, the non-native control group made fewer errors when rating stimuli items in the *illegal in 1* condition than those in the *low probability* condition.

However, the only form of sensitivity to phonotactic legality of English which the non-native speakers managed to show in the Lexical Decision Task is that both groups rejected stimulus items in the *illegal in 1* condition faster and with fewer errors than those in the *illegal in 2* condition. This result was replicated in the Non-word Rating Task as both non-native groups judged stimuli items in the *illegal in 1* condition to be less English-like than those in the *illegal in 2* condition. Recall that, in addition to being illegal in Arabic, stimuli items in the *illegal in 2* condition are also illegal in English as those in the *illegal in 1* condition. Therefore, Non-native speakers were actually predicted[56] to reject stimuli items in the *illegal in 2* condition faster or at least as fast as those in the *illegal in 1* condition.

In an attempt to explain this unpredicted result in the Non-word Rating Task, I made two assumptions. First, I assumed that non-native speakers may have managed to judge these items based on their knowledge of English only, in other words switching to a 'monolingual' English mode. Another assumption was that, influenced by their L1 phonotactic constraints, our non-native listeners did not correctly identify the non-L1 consonant clusters in the *illegal in 2* condition (e.g. /sr/ in /srud/). Non-native listeners may have misperceived the illegal cluster /sr/ as the legal one [ʃr] rendering it legal in English as well. On the other hand, their knowledge of English phonotactic constraints may not be subtle enough to cause such misidentification in the *illegal in 1* condition. This could explain why non-native listeners rejected stimuli items in the *illegal in 2* condition more slowly and with more errors than those in the *illegal in 1* condition.

---

[56] Although, as discussed in Chapter 4, null hypotheses were taken regarding non-native speakers' performance.

Because this pattern was replicated in the Lexical Decision Task, it seems that the latter proposal is more plausible. This is because of the fast processing the Lexical Decision Task entails. Non-native speakers are unlikely to have managed to use only their knowledge of English. Findings from Altenberg and Cairns (1983) in a visual presentation of the Lexical Decision Task lend support to the second proposal. Recall that when they asked English monolinguals and English-German bilinguals to rate how English-like non-words starting with consonant clusters illegal only in German (e.g. smatt), both groups' ratings were only influenced by the phonotactic constraints of English (see also Altenberg 2005b). However, when the same stimuli items where presented in a visual Lexical Decision Task, processing times of the bilinguals were affected by the phonotactic constraints of German as well.

The reason why our subjects in the lexical Decision Task did not show a clear effect of phonotactic legality as that shown in Altenberg and Cairns' Lexical Decision Task is that the items in our task were presented aurally while theirs were presented visually. Therefore, misperception of the initial illegal cluster as a result of phonotactic knowledge is likely to happen when items are only presented aurally rather than visually. Actually, this was the reason why they opted for a visual presentation of Lexical Decision Task, which is "in order to avoid the problem of subjects misperceiving illegal sequences" (ibid: 176).

Additionally, unlike in the auditory Lexical Decision Task where our native subjects failed to show an effect for non-L1 consonant clusters, in the visual Lexical Decision Task such an effect was shown. This indicates that, unlike in an auditory Lexical Decision Task, in a visual Lexical Decision Task decisions could be made based on the presence of an initial illegal cluster. In other words, while visual presentation of the non-word allows its classification as non-word earlier based on the initial illegal cluster, this does not happen when the non-word is presented aurally, in which case an evaluation of the whole item takes place.

Sensitivity to phonotactic probability. Let us first recap the results of native speakers regarding their sensitivity to the probability with respect to English phonotactics. In the Lexical Decision Task, native speakers decided that non-words with low phonotactic probability (/ðɔɪʒ/) are non-words more quickly and accurately than they did with those with high phonotactic probability. Native speakers were also faster and more accurate than non-native speakers in both groups in deciding that non-words were non-words in all conditions. These results confirm Hypothesis 6, which states that "Native speakers will reject stimuli items in the *low probability* condition faster and with fewer errors than those in the *high probability* condition" and Hypothesis 7, which states that "Native speakers will be faster in rejecting stimuli items and with fewer errors in all conditions than non-native speakers". These results lend support to previous findings that phonotactic probability not only affects subjective ratings but also processing times of stimuli in on-line tasks (Frisch et al. 2000; Vitevitch et al. 1997; Vitevitch and Luce 1998; 1999; Vitevitch and Luce 2005). They also show the difference between L1 speakers and L2 speakers concerning the speed and accuracy in accessing words in the lexicon. This suggests that one possible cause of non-native speakers' lack of fluency in speaking is their non-automatic access of words in their L2 lexicon. Recall that for non-native speakers the non-word stimuli mostly consisted of non-L1 phonemes and consequently influence from the L1 lexicon was unlikely to take place.

Despite lack of automaticity, the null Hypothesis 9, which states that "Non-native speakers will show no difference in RTs or error rates in rejecting stimuli items in the in the *high probability* condition and *low probability* condition" was rejected as our non-native speakers in the Lexical Decision Task have shown sensitivity to the phonotactic probability of English similar to that shown by native speakers. Our non-native subjects in both groups decided that non-words with low phonotactic probability (/ðɔɪʒ/) are non-words more quickly and accurately than they did with those with high phonotactic probability (e.g. /saɪv/). This sensitivity was replicated by the Non-word Rating Task as our non-native subjects in both groups judged stimuli items in the *high probability* condition more English-like than those in the *low probability* condition. In addition, the non-native experimental group made more errors when judging stimuli items in the *high probability* condition than in the *low probability* condition (i.e. they wrongly classified some *high probability* non-words as real words). Therefore, non-native speakers' sensitivity to the phonotactic probability shown by both the Non-word Rating and the Lexical Decision Tasks supports the proposal that phonotactic probability has a significant effect that is beyond and independent of neighbourhood density. The results also indicate that phonotactic probability effects are the result of abstract phonological information in memory that is independent of specific lexical representations of sound patterns. Another piece of indirect support for this proposal comes from subjects' (both native and non-native) sensitivity to the phonotactic legality of English shown by the Non-word Rating Task. Here sensitivity can not be the result of the presence of these constraints in lexical items; rather it is the result of the absence of these sequences (illegal clusters) in lexical items, which is more suggestive of the availability of an abstract knowledge of phonotactic constraints.

Taken together, results from the Non-word Rating and the Lexical Decision Tasks have shown that our EFL learners are sensitive to both the phonotactic legality and phonotactic probability of English. These findings suggest that like L1 infants, adult FL learners still possess the ability to detect the statistical regularities of phonemes in the little FL input they receive. Despite the fact that our EFL learners had never been taught English phonotactic constraints, they showed similar subtle knowledge of these constraints in the pre-test Non-word Rating Task. In addition, they showed subtle knowledge of phonotactic probability of English in both the pre-test Non-word Rating and the Lexical Decision Tasks despite getting little input and the fact that probabilistic phonotactics was not, and in fact, can not be taught.

It seems that learners' ability to detect statistical regularities of phonemes in the input is not confined to real language input. Recall, as discussed in Chapter 3 (Section 3.3.4), that Chambers et al. (2003) familiarized eight 16.5 month old infants to lists of CVC syllables which followed artificial phonotactic constraints (e.g. /b/ is always an onset and /p/ is always a coda). When infants were tested using the head-turn preference protocol using novel syllables that either followed the phonotactic constraints of the experiment or violated them, it was found that infants listened longer to novel syllables which violated the phonotactic constraints of the experiment, suggesting that infants acquired the phonotactic constraints of the experiments during training. Similar results, as discussed in Chapter 3, were also found with adult English speakers (Onishi et al. 2002).

Therefore, it seems that just like L1-learning infants (Jusczyk et al. 1993b), the results in the present study show that EFL learners are able to explore the phonotactic regularities of English from a relatively small amount of input and develop sensitivity to these phonotactics. In addition, as discussed above, these findings from EFL learners are, like those from infants, more suggestive that probability effects are the result of abstract phonological information in memory that is independent of lexical representations of sound patterns. However, it should be clear here that the word *abstract* here does not mean that this knowledge is universal. Rather the claim here is that phonotactic knowledge is language specific knowledge regarding the arrangement of phonetic segments and sequences of segments. This knowledge is acquired through exposure to EFL words and then independently stored in memory regardless whether if the exact lexical items have been stored in the lexicon or not. Therefore, our evidence from EFL learners shows that phonotactic effects are the result of drawing upon this independently stored knowledge rather than the result of direct analogy to existing words in the lexicon.

5.2.3 Results of the Word Spotting Task

RTs and error rates were counted and analysed. The researcher had to listen to the tape of each subject, write down the target words that were pronounced correctly and go back to the DMDX data file to check if there was a corresponding RT for that target word. RTs included in the analysis were only those accompanied by correct oral productions of the target words. Practice items and real-word fillers were excluded from analysis. On the other hand, an error was counted when there was no response or a response was either not accompanied by a correct oral production or accompanied by a wrong one. Other cases where an error was counted also include a) when the target word recorded on the tape was not intelligible and b) when the target word was pronounced correctly but there was no RT recorded because the subject reacted (i.e. pressed the response button) too late.

Recall that the aim of this task was to find if our subjects used the phonotactic constraints of English and Arabic in the lexical segmentation of English. Following the analysis by Weber and Cutler (2006), the best way to answer our research question is by comparing RTs and error rates of three boundary conditions (Common (e.g. /viːʃlain/), English (e.g. /viːdlain/) and Arabic (e.g. /viːblain/) to those in the *No boundary* condition (e.g. /viːflain/). If our subjects use the phonotactic constraints in each of the three boundary conditions in the lexical segmentation of English, their RTs and error rates in each of these condition are expected to be lower than those in the *No boundary* condition. In the following discussion, effects are considered statistically significant at $p \leq 0.05$ and as marginally significant where $0.05 < p \leq 0.1$. (a) Mean RTs in milliseconds measured from target word offset and (b) mean percentage of error rate for the three subjects groups (native control, non-native control and non-native experimental) in the four boundary conditions (Common boundary, English boundary, Arabic boundary and No boundary) are shown in Table 5.2 below.

Table 5.2 Mean RTs in milliseconds measured from word offset and mean percentage of errors in the pre-test for native control, non-native control and non-native experimental groups in four different boundary conditions.

Boundary type	Common Boundary	English Boundary	Arabic Boundary	No Boundary
(a) RTs in ms				
Native speakers	569	697	944	937
Non-native control	1114	1281	1245	1467
Non-native experimental	1137	1246	1283	1398
(b) % of Errors				
Native speakers	11	15.7	31.5	33.3

Non-native control	23.3	39.4	33.9	60
Non-native experimental	26.6	37.2	37.2	52.2

5.2.3.1 Native control

The first procedure was to find out if our native control group showed effects of different boundary conditions. Their mean RTs in milliseconds measured from target word offset and mean percentage of error rate for each condition are compared in Figure 5.13 and Figure 5.14 respectively.

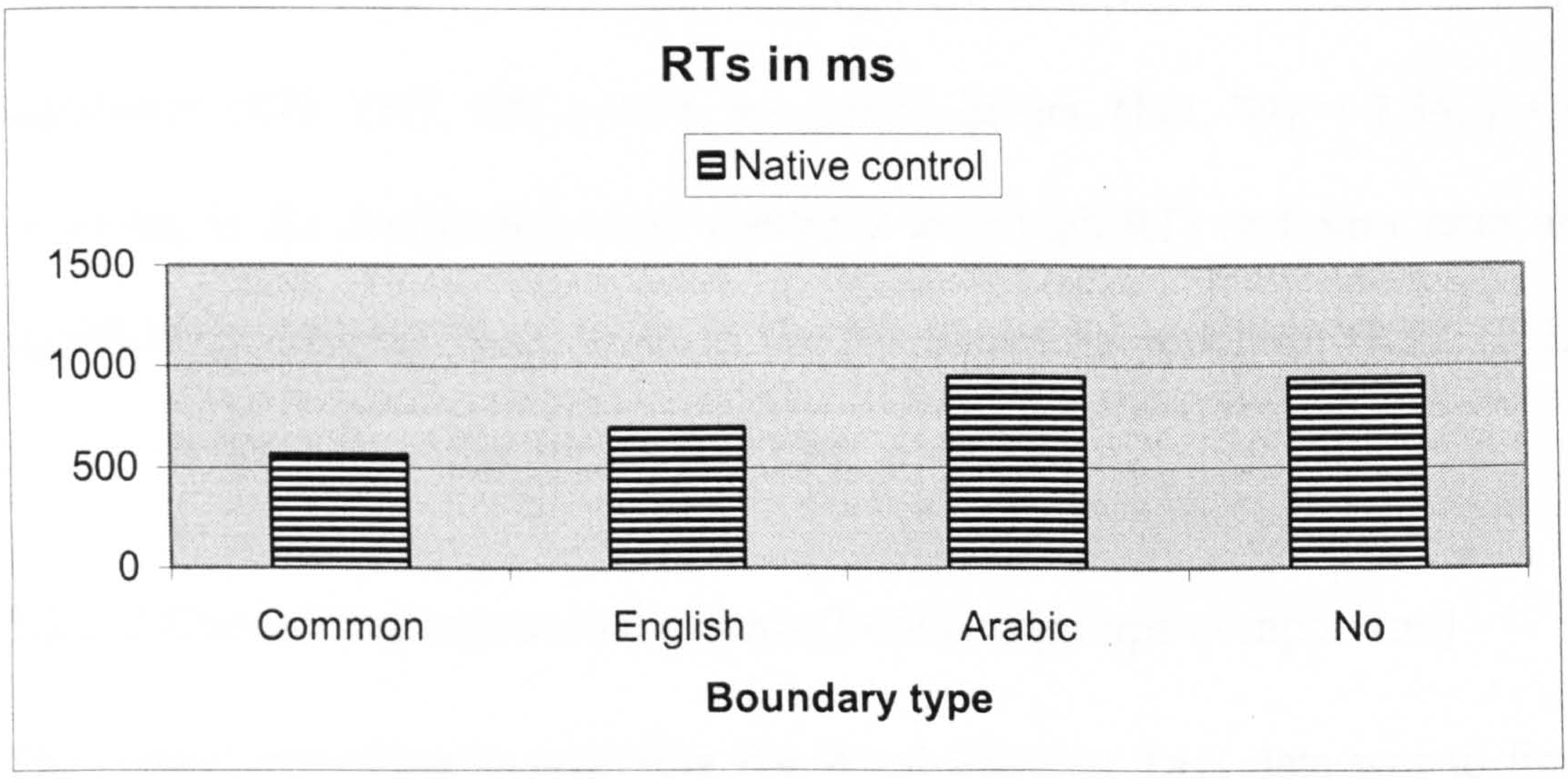

Figure 5.13 Native speakers' mean RTs in ms measured from target word offset in the WS Task in four different boundary conditions.

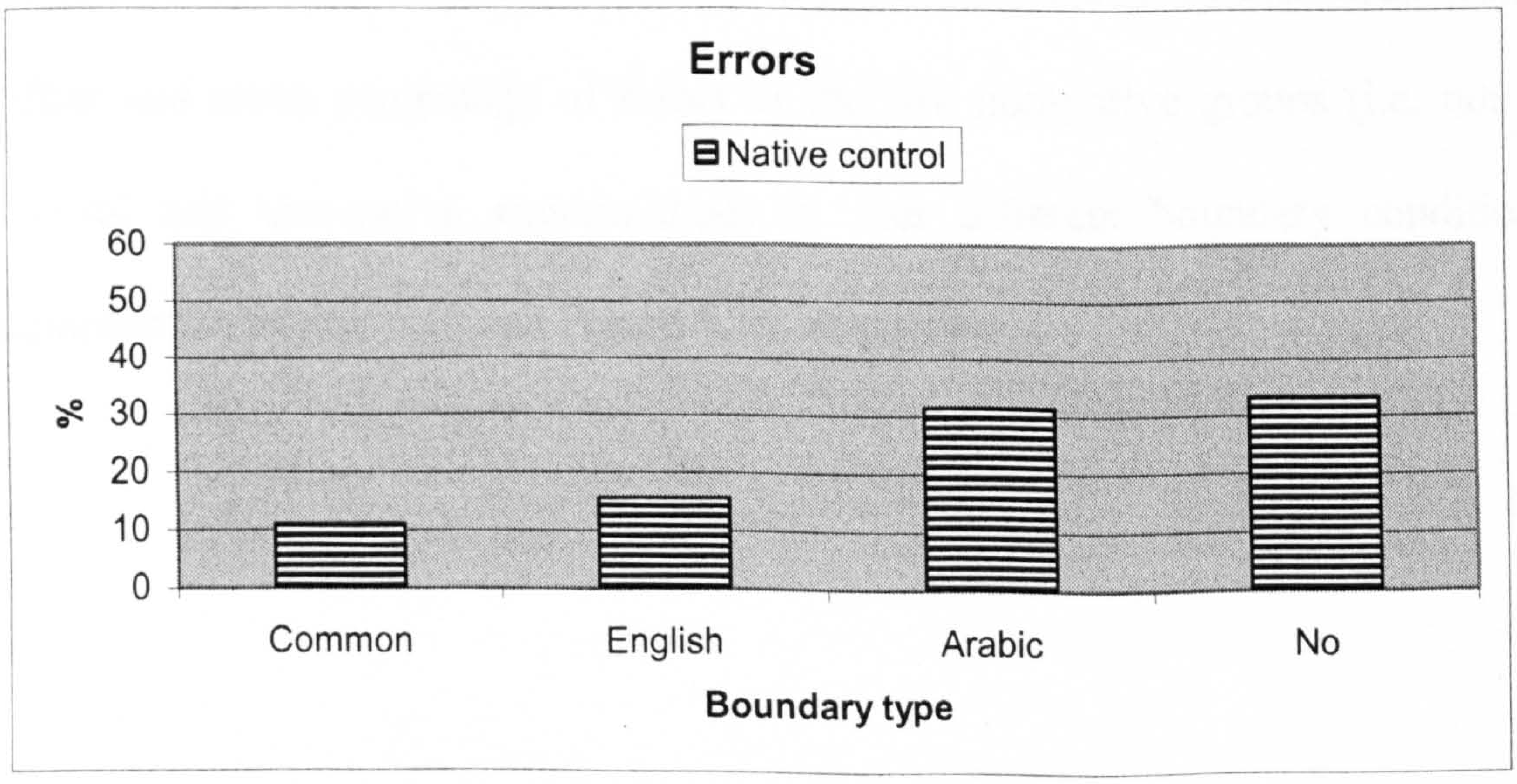

Figure 5.14 Native speakers' mean percentage of errors in the WS Task in four different boundary conditions.

Two separate one-way ANOVAs were conducted to compare native speakers' mean RTs (in the first ANOVA) and error rate (in the second) in the *Common, English and Arabic Boundary* conditions to those in the *No Boundary* condition. Post hoc test results showed that native English speakers were significantly faster and more accurate in spotting words in the *Common Boundary* condition than those in the *No Boundary* condition (RTs, (F(3, 44) = 3.31, p= 0.005; Errors, (F(3, 44) = 2.84, p= 0.034). They were also faster and more accurate in the *English boundary* condition than in the *no boundary* condition although the difference in RTs was marginally significant (RTs, (F(3, 44) = 3.31, p= 0.065; Errors, (F(3, 44) = 2.84, p= 0.047). However, in the *Arabic Boundary* condition their high RTs and error rates were not significantly different from those in the *No Boundary* condition (RTs, (F(3, 44) = 3.31, p = 0.846; error rate, (F(3, 44) = 2.84, p = 0.828).

5.2.3.2 Comparing non-native groups (between groups comparison)

The second procedure in analysing the Word Spotting Task data was to find out if there was any significant difference between the RTs and errors of the two non-native groups in the four different boundary conditions. Mean RTs measured from item offset and mean percentage of errors of the two non-native groups (i.e. non-native control and non-native experimental) in four different boundary conditions are compared in Figure 5.15 and Figure 5.16, respectively.

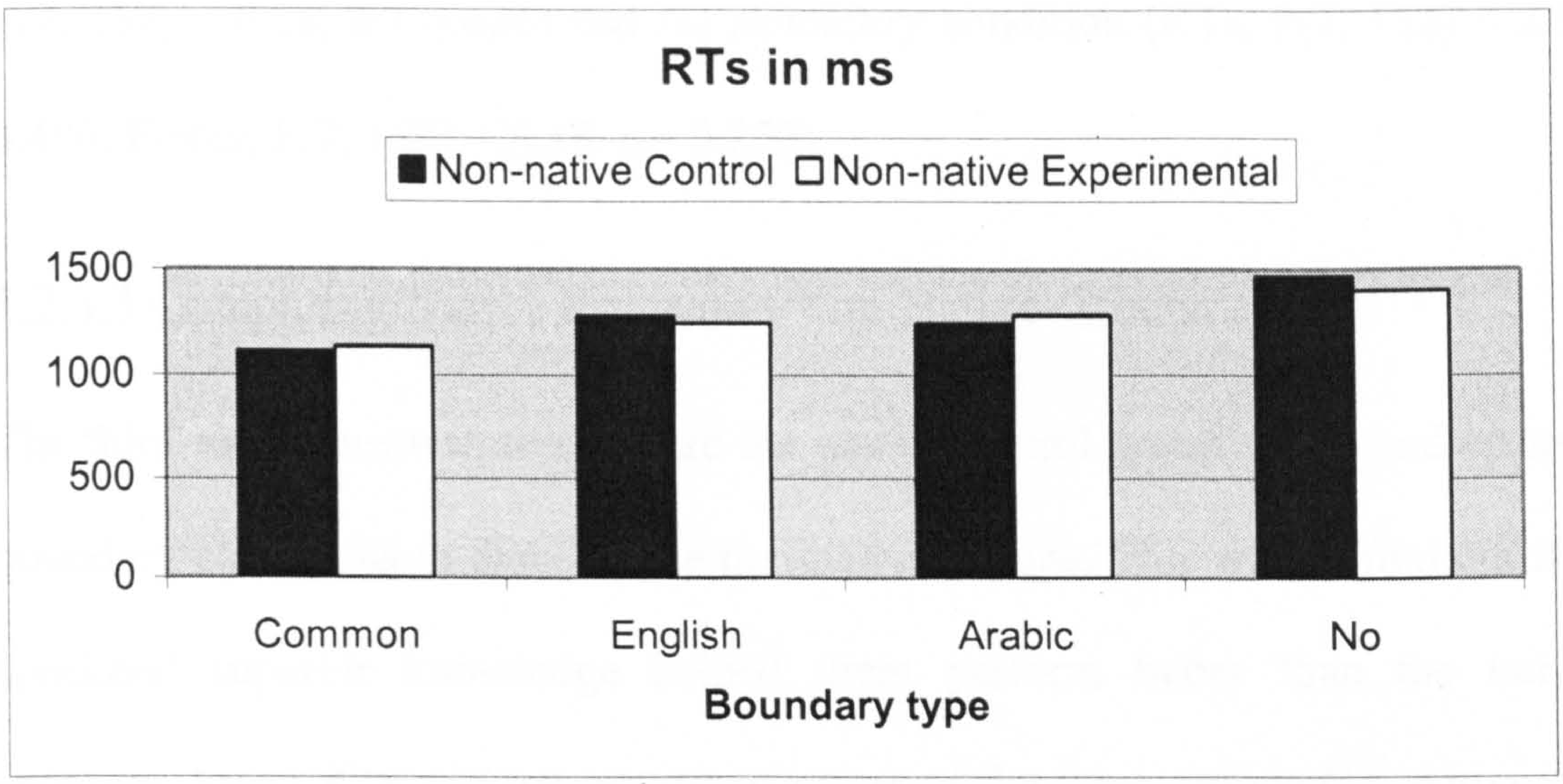

Figure 5.15 Mean RTs measured from target word offset for the two non-native groups in the Word Spotting Task in four different boundary conditions.

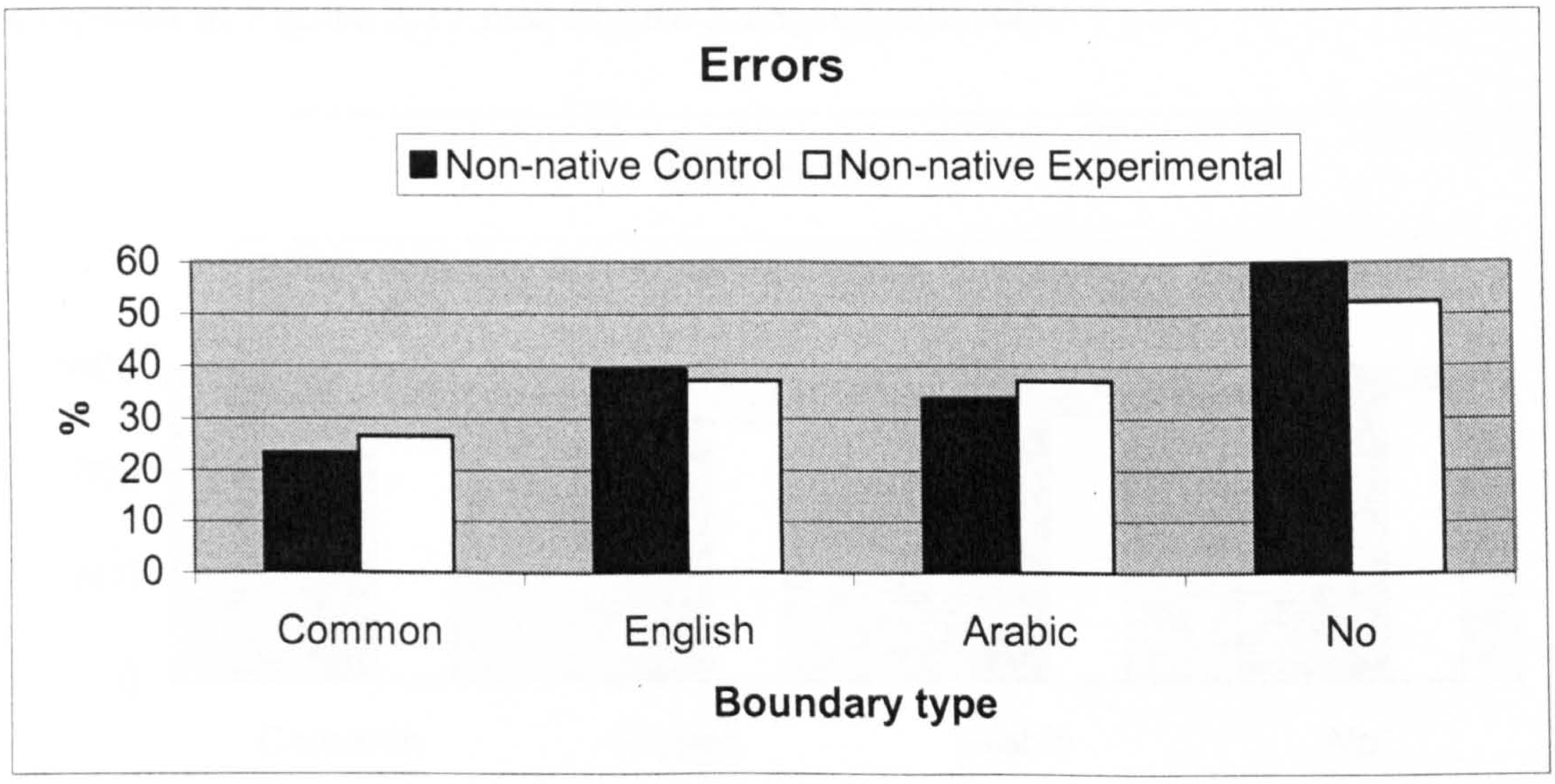

Figure 5.16 Mean percentage of errors of the two non-native groups in the Word Spotting Task in four different boundary conditions.

Two separate one-way ANOVAs were conducted to compare mean RTs (in the first ANOVA) and error rate (in the second) of the two non-native groups in all four conditions. Post hoc test results showed that the two groups had statistically indistinguishable RTs and error rates in all four conditions; *Common Boundary* condition (RTs, F(7, 152) = 2.93, *p*= 0.820; Errors, F(7, 152) = 6.18, *p*= 0.628), *English Boundary* condition (RTs, F(7, 152) = 2.93, *p*= 0.723; Errors, F(7, 152) = 6.18, *p*= 0.746), *Arabic Boundary* condition (RTs, F(7, 152) = 2.93, *p*= 0.708; Errors, F(7, 152) = 6.18, p= 0.628) and *No Boundary* condition (RTs, F(7, 152) = 2.93, p= 0.490; Errors, F(7, 152) = 6.18, p= 0.259).

5.2.3.3 Comparing Native Control to Non-Native Groups

The third procedure was to compare the native control group's RTs and errors in all boundary conditions to those of the non-native groups. This was to find out if native speakers' superior knowledge helped them perform better than the non-native speakers. Mean RTs and percentage of errors of the three groups (i.e. native control, non-native control and non-native experimental) in the four boundary conditions are compared in Figure 5.17 and Figure 5.18, respectively.

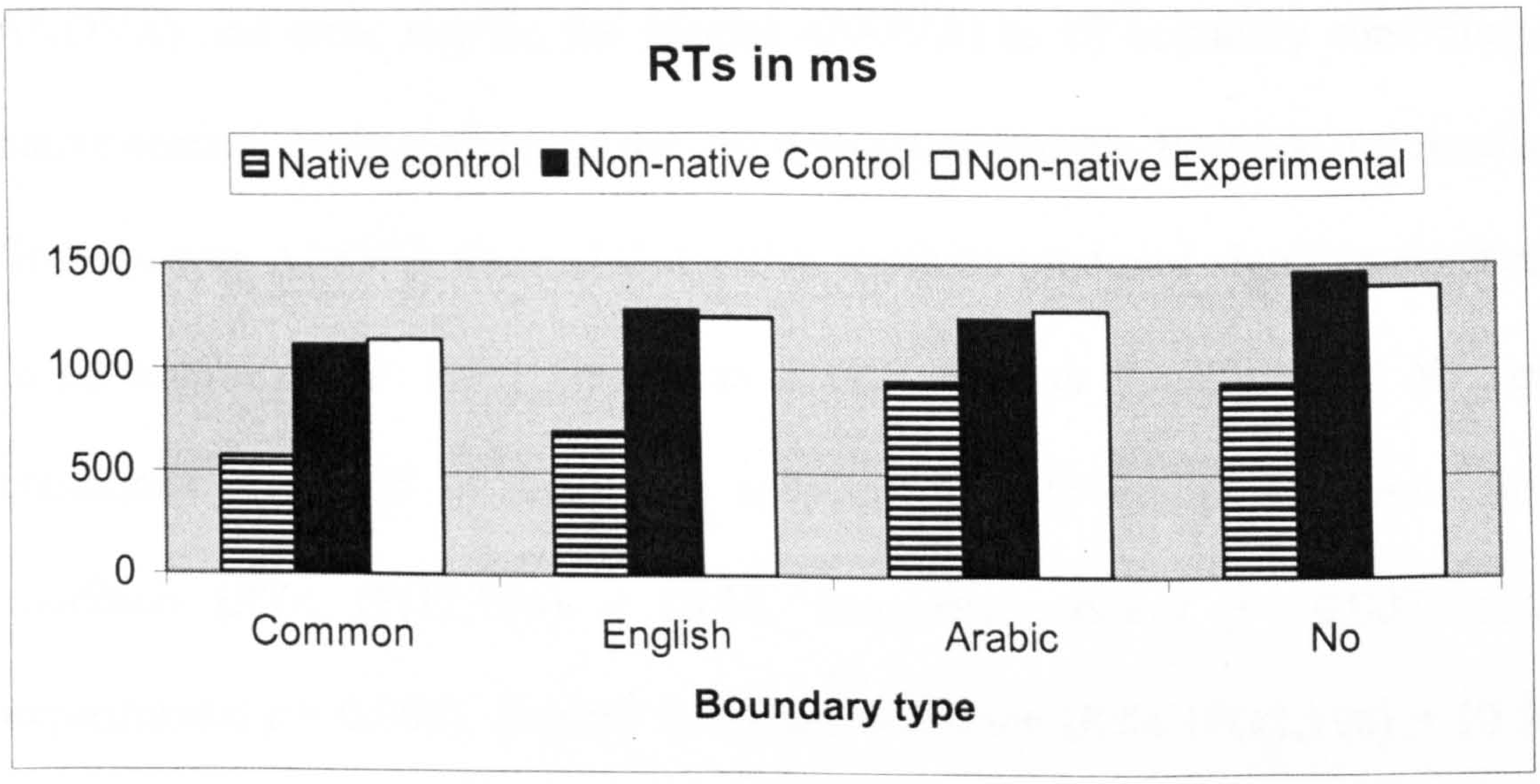

Figure 5.17 Mean RTs in ms measured from word offset in the Word Spotting Task for the three subject groups (native control, non-native control and non-native experimental) in four different boundary conditions.

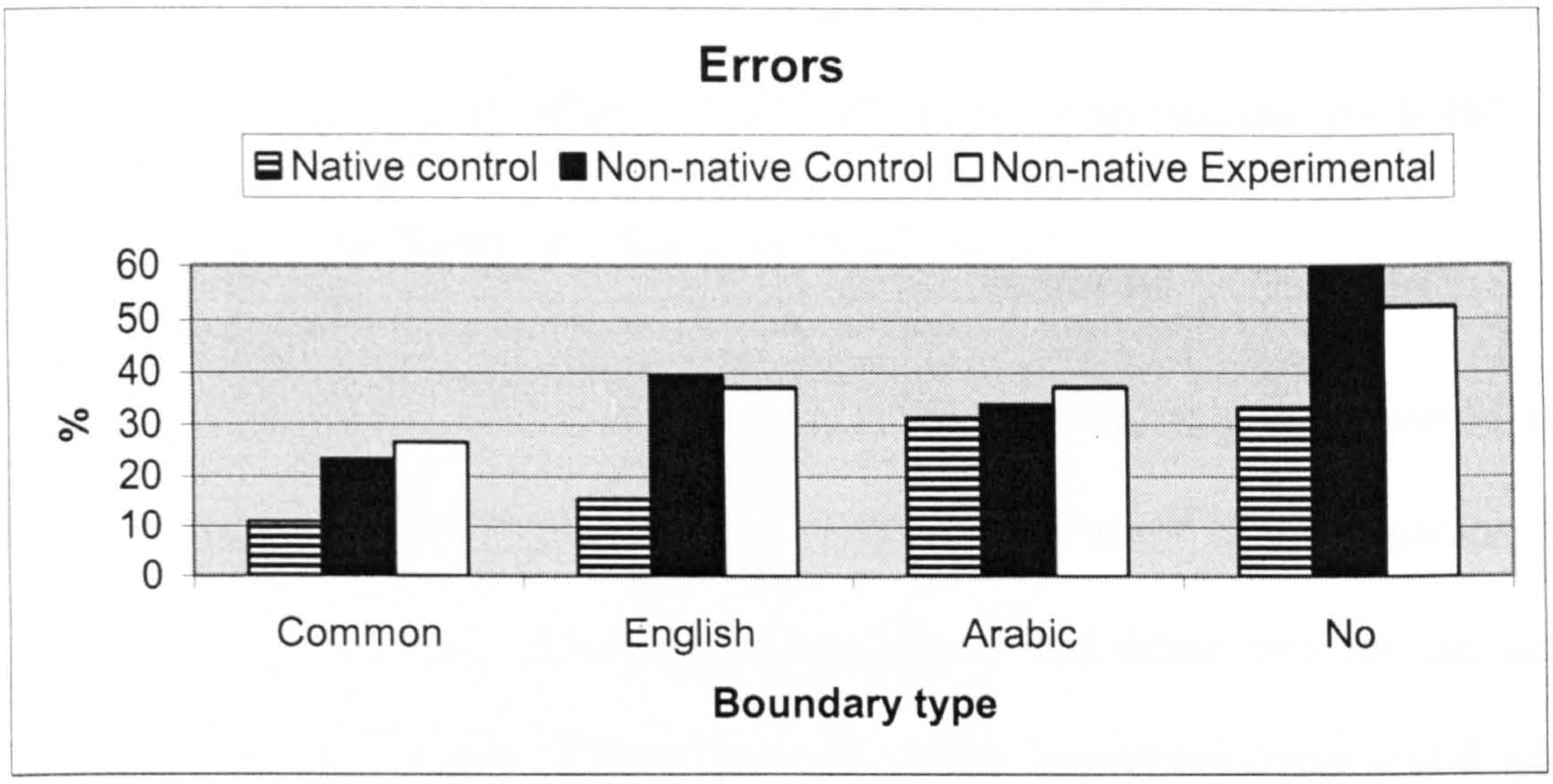

Figure 5.18 mean percentage of errors in the Word Spotting Task for the three subject groups (native control, non-native control and non-native experimental) in four different boundary conditions.

Two separate one-way ANOVAs were conducted to compare mean RTs (in the first ANOVA) and error rate (in the second ANOVA) in all boundary conditions of the native control group to those of the two non-native groups. Post hoc test results of the first one-way ANOVA showed that native speakers produced significantly faster RTs in all conditions but lower error rates in only *English Boundary* and *No Boundary* conditions compared to non-native speakers in both groups (*Common Boundary* condition (RTs, (F(11,196) = 10.56, non-native control *p*= 0.000; non-native experimental *p*= 0.000), *English Boundary* condition (RTs, (F(11,196) = 10.56, non-native control *p*= 0.000; non-native experimental *p*= 0.000; Errors, (F(11, 196) = 6.74, non-native control *p*= 0.003; non-native experimental *p*= 0.007), *Arabic Boundary* condition (RTs, (F(11,196) = 10.56, non-native control *p*= 0.010; non-native experimental *p*= 0.004) and *No Boundary* condition (RTs, (F(11,196) = 10.56, non-native control *p*= 0.000; non-native experimental *p*= 0.000; Errors, (F(11, 196) = 6.74, non-native control *p*= 0.001; non-native experimental *p*= 0.017). In the *Common Boundary* and *Arabic Boundary* conditions, native speakers produced statistically indistinguishable error rates from those of the non-native groups (*Common Boundary* condition (Errors, (F(11, 196) = 6.74, non-native control *p*=

0.122; non-native experimental $p= 0.060$, *Arabic Boundary* condition (Errors, (F(11, 196) = 6.74, non-native control $p= 0.760$; non-native experimental $p= 0.467$).

5.2.3.4 Non-native groups (within groups comparison)

The fourth procedure was to investigate if the non-native groups would show the same effects of different condition as native speakers.[57] Four separate One-way ANOVAs (two for the non-native control group and other two for the non-native experimental group) were conducted to compare non-native control and non-native experimental groups' mean RTs (in the first two ANOVAs) and error rate (in the other two ANOVAs) in the *Common, English and Arabic Boundary* conditions to those in the *No Boundary* condition. Post hoc test's results showed that both groups were faster and more accurate in the *Common Boundary* condition than in the *No Boundary* condition (RTs, non-native control $p= 0.000$; non-native experimental $p= 0.014$; Errors, non-native control $p= 0.000$; non-native experimental $p= 0.001$) Interestingly, although not faster, in the *English Boundary* condition they were more accurate than in the *No boundary* condition (RTs, non-native control $p= 0.065$; non-native experimental $p= 0.145$; Errors, non-native control $p= 0.000$; non-native experimental $p= 0.045$).[58] Lastly, in the *Arabic Boundary* condition, both groups were significantly more accurate than in the *No Boundary* condition (Errors, non-native control $p= 0.000$; non-native experimental $p= 0.045$). However, only the non-native control group was significantly faster $p = 0.027$ in this condition. This latter result could be attributed to the fact that the non-native experimental group produced, although not significantly, higher RTs in the *Arabic Boundary* condition (mean =

---

[57] Recall that null hypotheses were set regarding non-native speakers performance in the Word Spotting task.

[58] That is interesting because it shows that the non-native groups started using English phonotactic constraints to some extent at pre-test in segmentation.

1283ms) and lower RTs in the *No Boundary* condition (mean = 1398ms) than the non-native control group (means 1245ms and 1467ms respectively) thus reducing the difference between the *Arabic Boundary* and *No Boundary* conditions.

5.2.3.5 Discussion

Recall that unlike the Non-word Rating and Lexical Decision Tasks which investigated subjects' sensitivity to the phonotactic probability of English, the current Word Spotting Task did not attempt such investigation. Rather, it only tried to investigate research question 3, namely if Arabic speakers EFL learners used the phonotactic constraints of Arabic and English in the lexical segmentation of running speech in English. However, we first need to recap results from our native speakers to ensure that this task is controlled and reliable.

Results from our native speakers of American English lend support to previous findings discussed in Chapter 3 (e.g. McQueen 1998; Weber 2000; Weber and Cutler 2006), that L1 listeners use their native language phonotactic constraints in lexical segmentation. The native speakers in the present study were faster and more accurate in detecting a word aligned with a boundary constraint common to English and Arabic (e.g. ʒi:tlɔ:rd) than when it was not aligned (ʒi:flɔ:rd). They were also faster and more accurate in detecting a word aligned with an English boundary constraint (e.g. ʒi:dlɔ:rd) than when it was not. However, they were as fast and accurate in detecting words in the *Arabic Boundary* and *No Boundary* conditions. Moreover, native subjects were faster than EFL learners in detecting words in all conditions. But, they were only more accurate in detecting words in the *English* and *No Boundary* conditions.

That native English speakers were overall faster than non-native speakers suggests that a difficulty for EFL learners in listening may be non-automaticity of lexical segmentation ability and bottom-up word recognition skills. These findings confirm Hypothesis 10, which states that "Native speakers will spot words in the *Common Boundary* and *English Boundary* conditions faster and with fewer errors than those in the *No Boundary* conditions" and Hypothesis 11, which states that "Native speakers will show no difference in spotting words in the *Arabic Boundary* and *No Boundary* condition". The findings also partially confirm Hypothesis 12, which states that "native speakers will be faster and more accurate than non-native speakers in spotting words in all conditions", as native speakers were not more accurate than non-native speakers in spotting words in the *Common Boundary* and *Arabic Boundary* conditions.

An interesting result but which is unrelated to the central issue in the current study was found. Our native English speakers, who have no knowledge of Arabic were faster and more accurate (although not significantly), in detecting a word aligned with a boundary constraint common to English and Arabic (e.g. ʒi:tlɔ:rd) than when it was aligned with a boundary constraint in English only (e.g. ʒi:dlɔ:rd) (mean RTs, 569ms vs. 697ms; mean percentage of errors, 11% vs. 15.7%, respectively). Upon examining Weber and Cutler's (2006) results, I found a similar trend.[59] Again, although not significantly, their native English speakers, who had no knowledge of German, were faster and more accurate in detecting a word aligned with a boundary constraint common to English and German than when it was aligned with a boundary constraint

---

[59] Weber and Cutler did not discuss this result.

in English only (e.g. in target words starting with /l/, mean RTs, 482ms vs. 543ms; mean percentage of errors, 11.9% vs. 17.2%, respectively. Other evidence that non-English phonotactic constraints can show effect on monolingual English speakers comes from Altenberg and Cairns (1983). In their visual Lexical Decision Task, they found that monolingual English speakers were faster in rejecting non-words that started with clusters illegal only in German than those that were legal in both languages.

These findings are interesting. Why should a monolingual be affected by non-L1 phonotactic constraints (i.e. illegal clusters)? One plausible explanation is that phonotactic constraints that are common to two or more languages are likely to be universal. Therefore, their effect on processing might be stronger than those illegal in the L1 only. However, this explanation may not be sufficient as monolingual English speakers in Altenberg and Cairns' study were affected by consonant clusters that were illegal only in German.

On the other hand, results from our non-native groups allow us to reject the null Hypothesis 13, which states that "Non-native speakers will show no difference in RTs or error rates in spotting words in the *No Boundary* condition on the one hand and the other three conditions on the other". These results confirm Weber and Cutler's findings that L2 listeners transfer their L1 phonotactic constraints and use them when segmenting speech in an L2. Our EFL learners were faster and more accurate in detecting a word aligned with an Arabic boundary constraint (e.g. ʒi:blɔ:rd) than when it was not. Recall that results from our Non-word Rating Task showed that non-native subjects used their knowledge of English phonotactic constraints when judging how English-like non-words sounded. In addition, our Lexical Decision Task failed to show that our non-native subjects were affected by their L1 phonotactic knowledge during on-line processing. Results from the current Word Spotting Task, however, suggest that non-native subjects use their L1 phonotactic constraints to guide their segmentation of running speech in an L2, suggesting that, unlike the case with off-line processing, operating in an L2 monolingual mode is difficult during on-line processing.

The most interesting pre-test result, however, is that our low-intermediate level EFL learners seem to have implicitly acquired some knowledge of English phonotactic constraints which they use when segmenting speech in English. They were more accurate in detecting a word aligned with an English boundary constraint (e.g. ʒi:dlɔ:rd) than when it was not. This is also confirmed by the fact that subjects were on average fastest and most accurate when Arabic and English phonotactic constraints joined forces in the *Common Boundary* condition.

Just like L1-learning infants (Jusczyk et al. 1993b), these results show that EFL learners are able to explore the phonotactic regularities of English based on a relatively small amount of input, as also confirmed by the Non-word Rating and Lexical Decision Tasks. Not only that but, as the Word Spotting Task has shown, it seems that our non-native subjects have started using this knowledge on-line in lexical segmentation as they were more accurate in detecting a word aligned with an English boundary constraint (e.g. ʒi:dlɔ:rd) than when it was not.

The problem for L2 learners, however, is that they have L1 phonotactic constraints in place. These constraints, in addition to the limited input non-immersion learners receive, may slow the acquisition of L2 phonotactic constraints. It is here where the role of awareness raising about these constraints and their use in lexical segmentation becomes a vital consideration. Given relatively little input and L1 influence, teaching phonotactics in an EFL context might provide a shortcut to using L2 English constraints in lexical segmentation by directing learners' attention to the presence of these cues. Post-test results below will show us if this type of teaching was effective in increasing subjects' sensitivity to the phonotactic constraints of English and improving their use of these cues in the lexical segmentation of running speech in English.

5.3 Post-tests' results: the effect of the treatment

5.3.1 Results of the Non-word Rating Task

The same analysis used in the pre-test was also used in the post-test. In this task two dependent variables were analysed, namely ratings and error rate. An error was counted when either a rating was not given for a non-word item or the item was wrongly rated as a real word. Practice items and real-word fillers were not included in the analysis. Recall that it was only the non-native experimental group who received the treatment. The treatment was aimed at teaching only phonotactic constraints of English and but not unteachable phonotactic probability. However, both groups received a considerable amount of input through in-class activities and the weekly tasks they were set. This type of naturalistic classroom input may have increased both groups' sensitivity to the phonotactic probability of English. Therefore, results relevant to both topics will be analysed.

5.3.1.1 Comparing non-native control pre-test with post-test results

The first procedure in analyzing the non-native control group's post-test results for the Non-word Rating Task was to find out if there is any significant difference between pre-test and post-test results in all four conditions. The non-native control group's mean ratings and mean percentage of errors in four different phonotactic conditions in the pre-test and post-test are compared in Figure 5.19 and Figure 5.20, respectively.

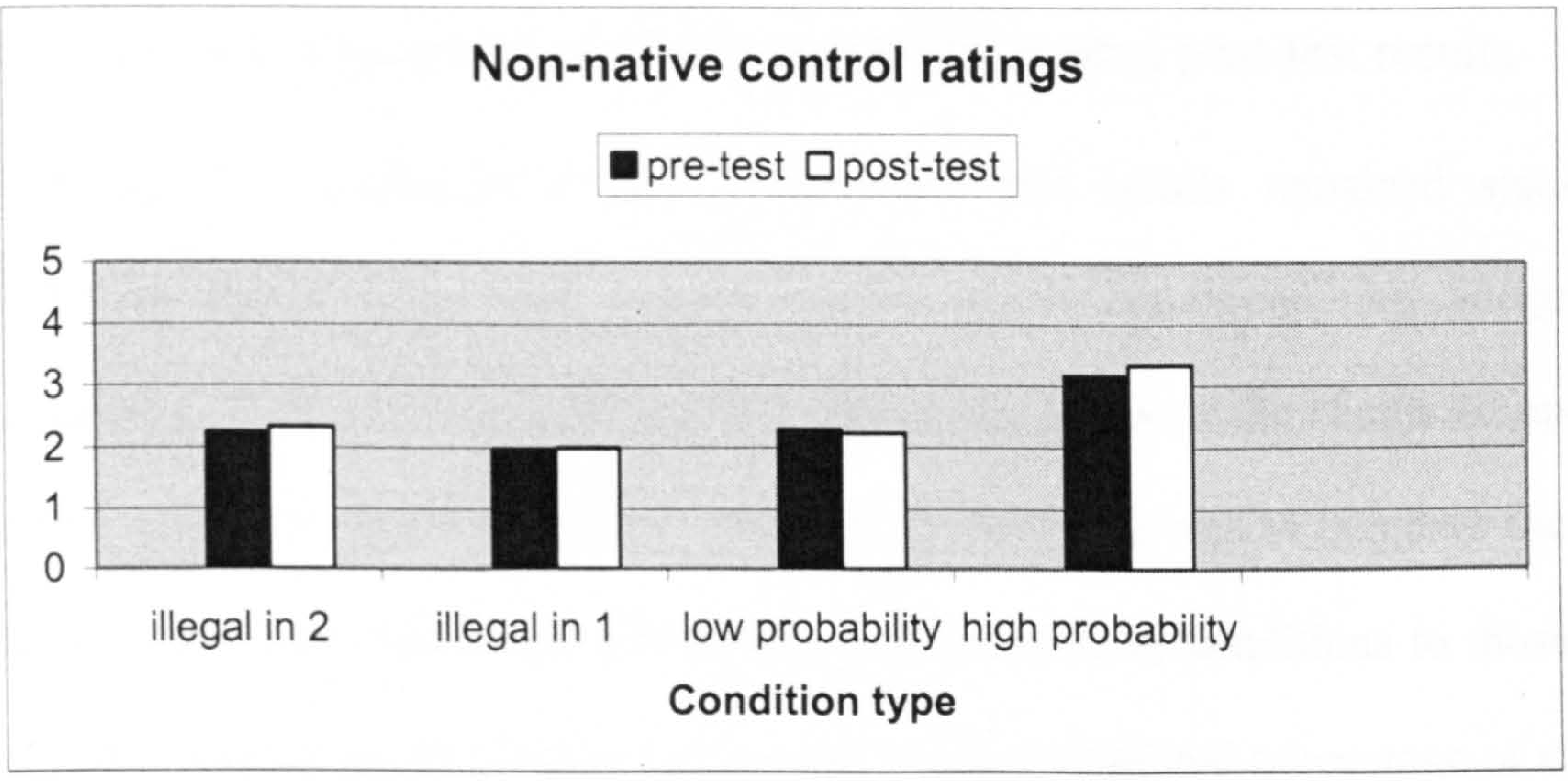

Figure 5.19 Non-native control group's mean ratings in the NWR Task in four different phonotactic conditions in the pre-test and post-test.

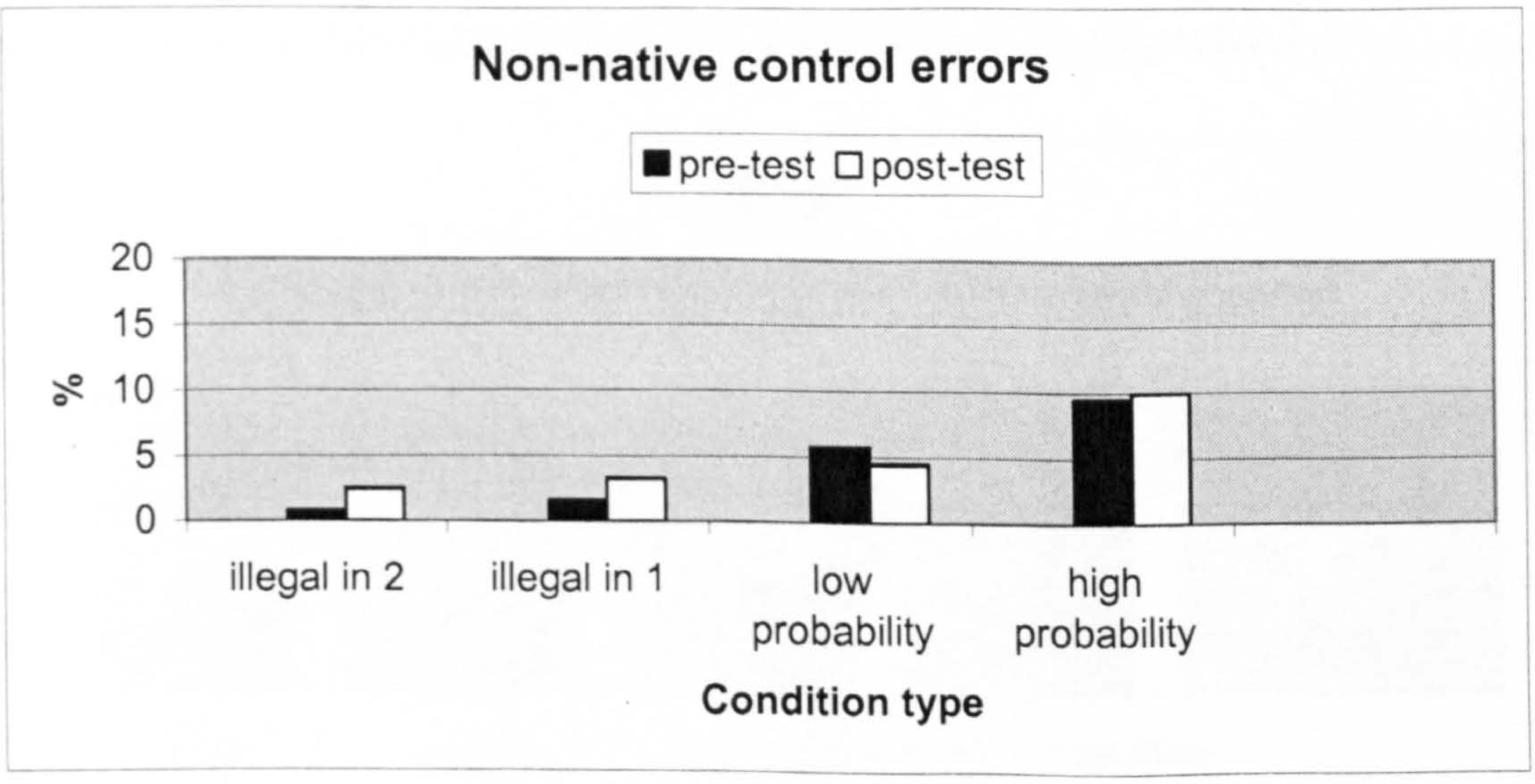

Figure 5.20 Non-native control group's mean percentage of errors in the NWR Task in four different phonotactic conditions in the pre-test and post-test.

Two one-way ANOVAs were conducted to compare the non-native control group's post-test ratings (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed the non-native control group's ratings and errors in the post-test in all conditions remained statistically indistinguishable from theirs in the pre-test (*illegal in 2* condition (Ratings, $p = 0.646$; Errors $p = 0.432$), *illegal in 1* condition (Ratings, $p = 0.959$; Errors $p = 0.432$), *the low probability* condition (Ratings, $p = 0.683$; Errors $p = 0.370$) and *the high probability* condition (Ratings, $p = 0.321$; Errors $p = 0.844$).

5.3.1.2 Comparing native control to non-native control post-test results

Although the non-native control group's post-test results remained statistically indistinguishable from their pre-test results in all conditions, any insignificant difference in the post-test results could change the status of the results compared to native control's results. Therefore, the second procedure was to compare the native English-speaking control group's ratings and errors in all conditions to those of the non-native control group's post-test results. Mean ratings and percentage of errors of the native control and non-native control's pre-test and post-test results in the four conditions are compared in Figure 5.21 and Figure 5.22, respectively.

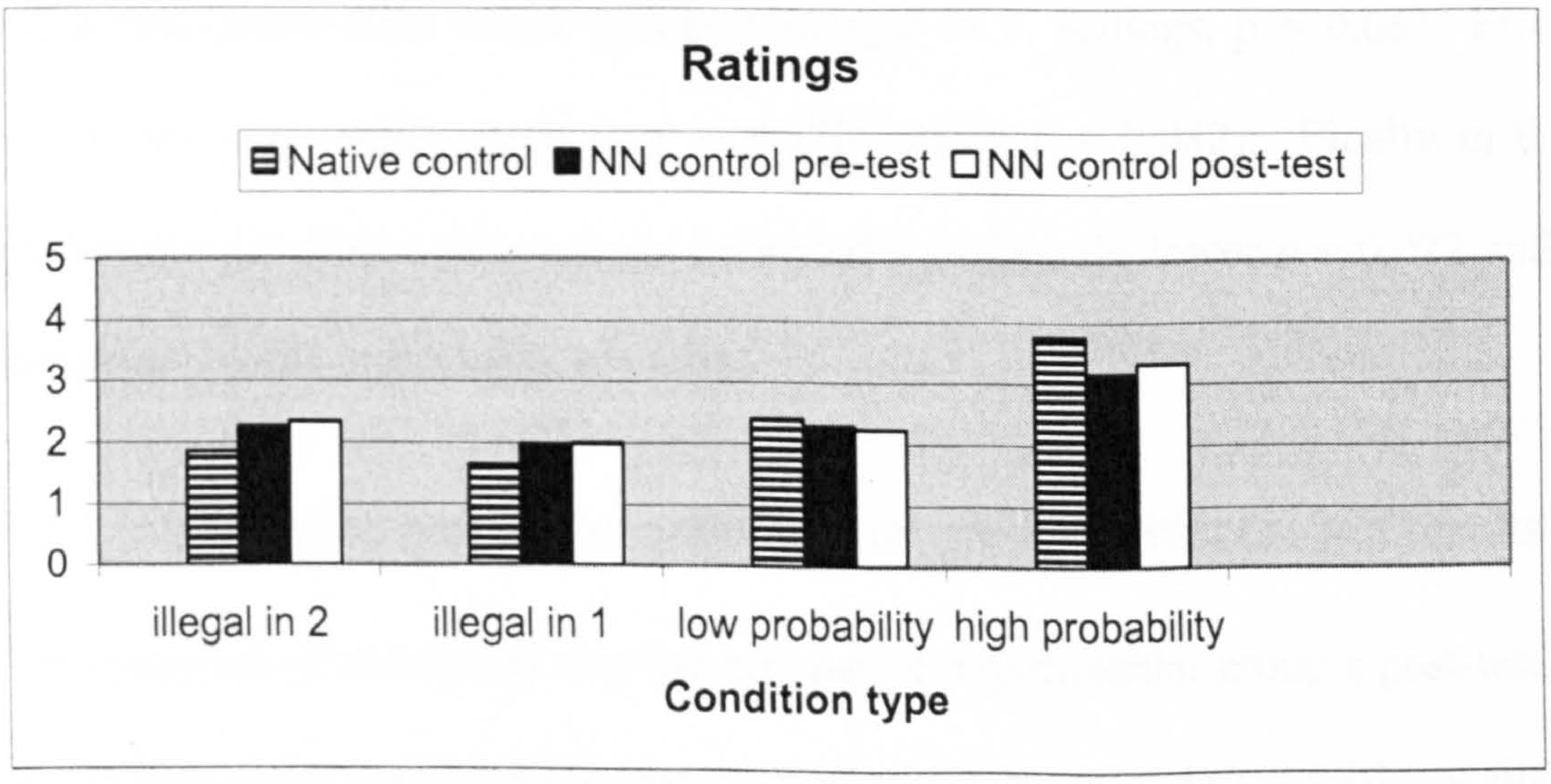

Figure 5.21 Native control group and non-native control group's pre-test and post-test ratings in the NWR Task in four different phonotactic conditions.

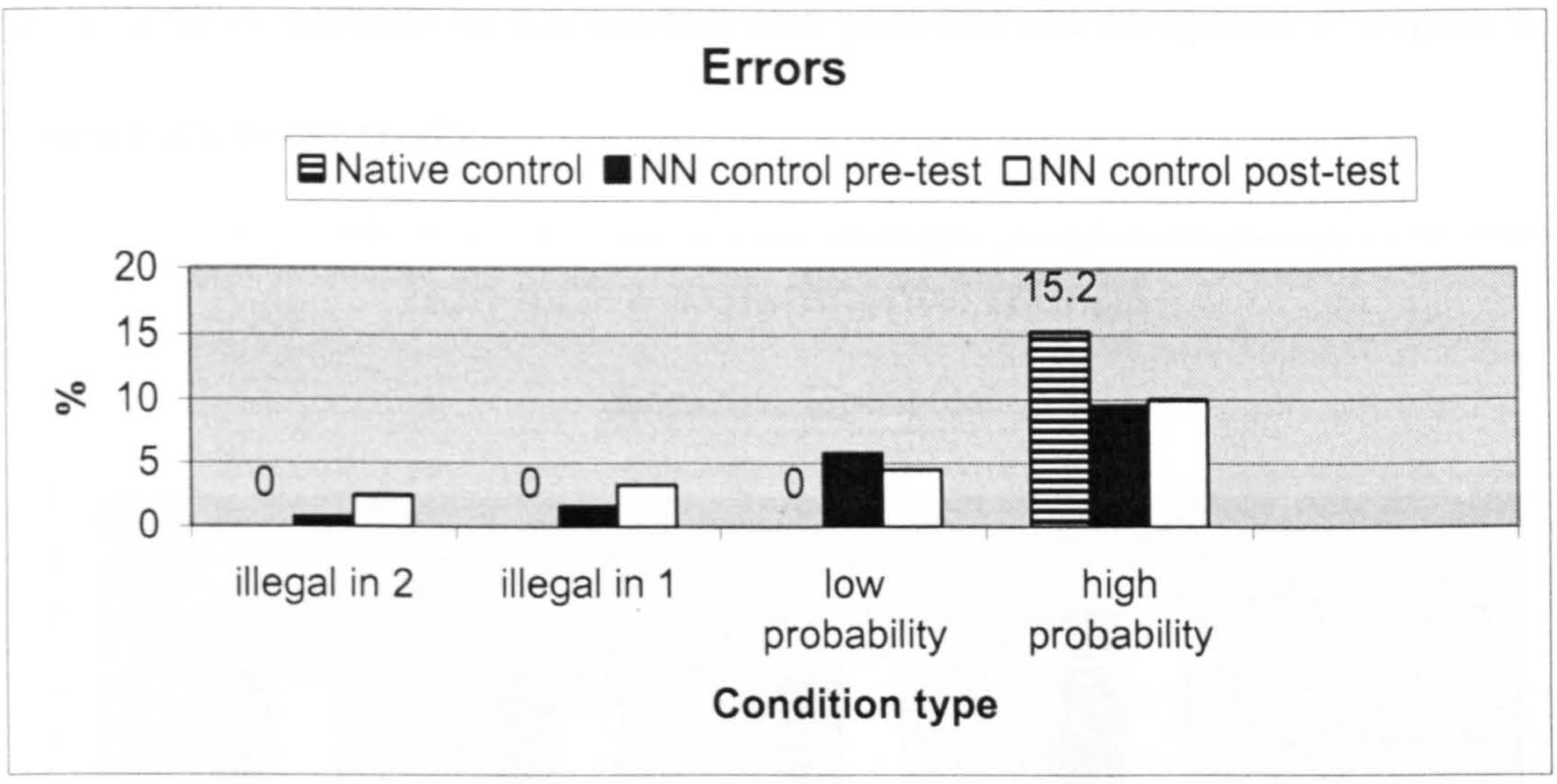

Figure 5.22 Native control group and non-native control group's pre-test and post-test errors in the NWR Task in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the native control group's ratings (in the first ANOVA) and error rate (in the second) to those of the non-native control group's post-test in all four conditions. Post hoc test results showed that the non-native control group's results did not change in the post-tests compared to native speakers'. In the *illegal in 2* condition their ratings remained higher $p = 0.012$ and their errors statistically indistinguishable $p = 0.308$ from native speakers. In the *illegal in 1* and *low probability* conditions both their ratings and errors remained statistically indistinguishable from native speakers (*illegal in 1,* Ratings, $p = 0.061$; Errors $p = 0.174$; *low probability,* Ratings, $p = 0.370$; Errors $p = 0.082$). Finally in the *high probability* condition, their ratings remained significantly lower $p = 0.022$ and errors statistically indistinguishable $p = 0.061$.

5.3.1.3 Comparing non-native experimental pre-test with post-test results

The first procedure in analyzing the non-native experimental group's post-test results of the Non-word Rating Task was to find out if there is any significant difference between pre-test and post-test results in all four conditions. The non-native experimental group's mean ratings and mean percentage of errors in four different phonotactic conditions in the pre-test and post-test are compared in Figure 5.23 and Figure 5.24, respectively.

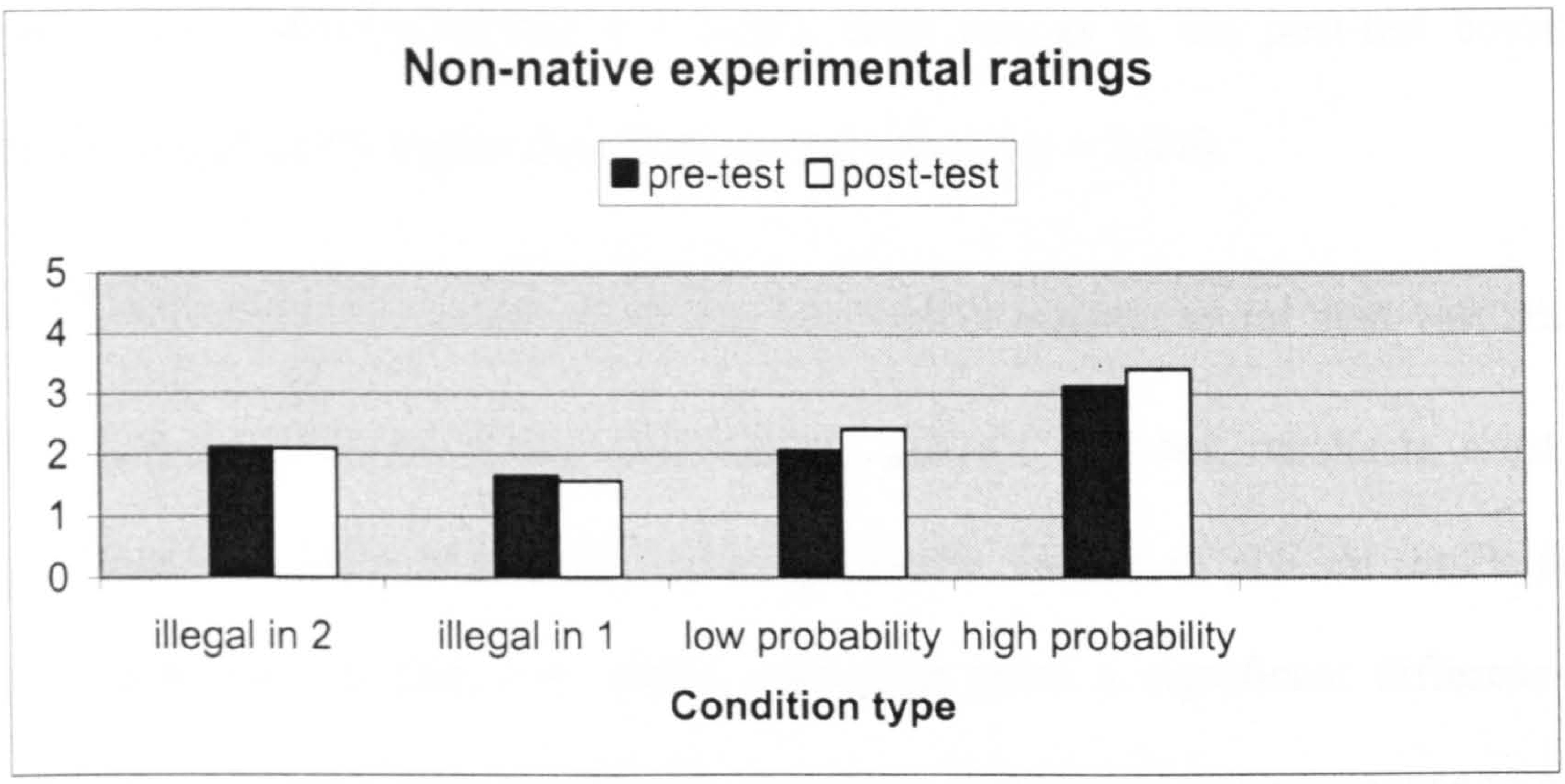

Figure 5.23 Non-native experimental group's mean ratings in the NWR Task in four different phonotactic conditions in the pre-test and post-test.

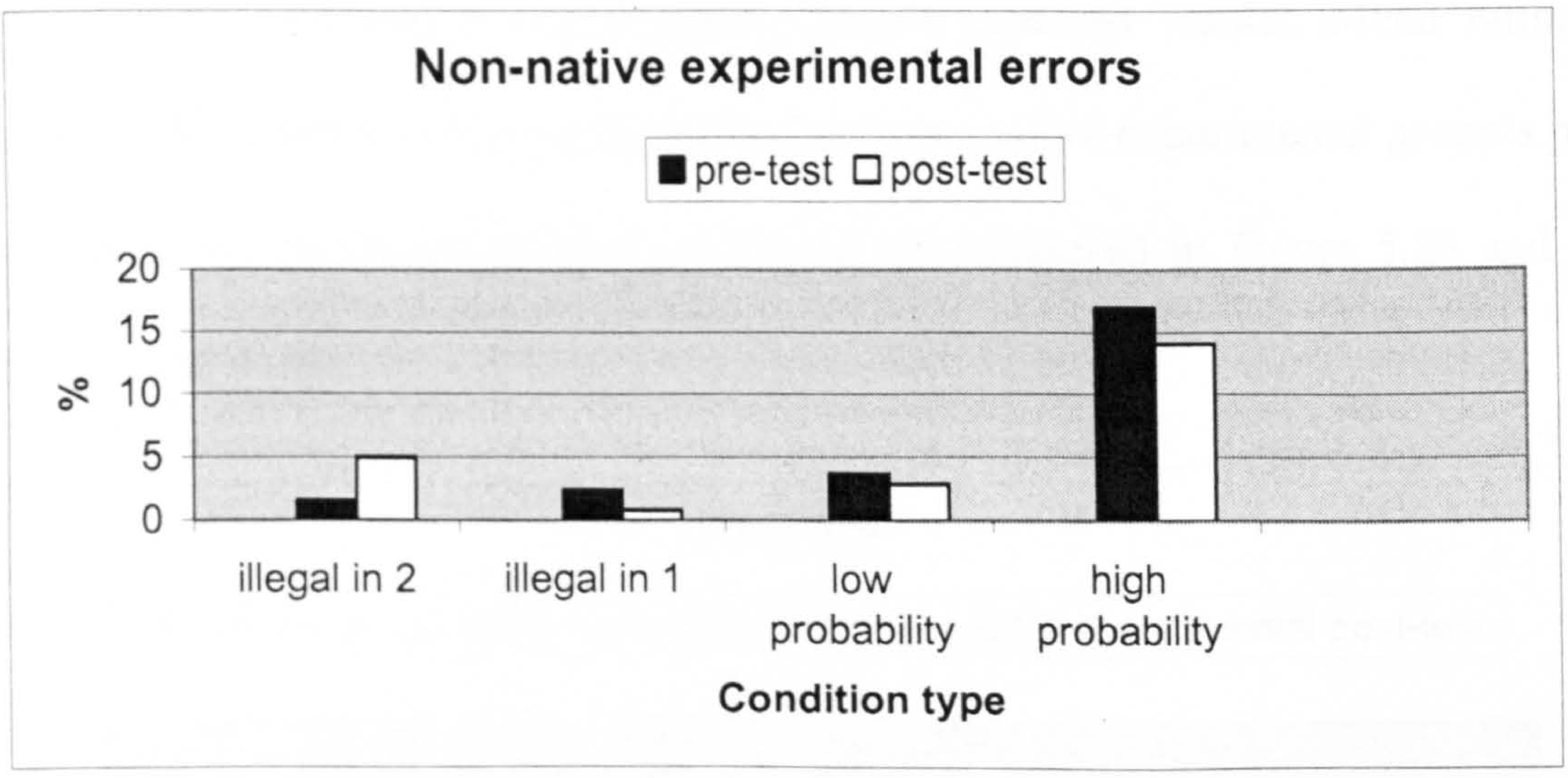

Figure 5.24 Non-native experimental group's mean percentage of errors in the NWR Task in four different phonotactic conditions in the pre-test and post-test.

Two one-way ANOVAs were conducted to compare the non-native experimental group's post-test's ratings (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed that the non-native experimental group's ratings and errors in the post-test remained statistically indistinguishable from theirs in the pre-test in three conditions: *illegal in 2* condition (Ratings, $p = 0.928$; Errors $p = 0.230$), *illegal in 1* condition (Ratings, $p = 0.683$;

Errors $p = 0.548$), and *the high probability* condition (Ratings, $p = 0.108$; Errors $p = 0.764$). However, in the *low probability* condition although the error rate remained statistically indistinguishable $p = 0.293$, their ratings in the post-test became just barely significantly higher than theirs in the pre-test $p = 0.049$.

5.3.1.4 Comparing native control to non-native experimental post-test results

The fact that the non-native experimental group's post-test results in most of the conditions remained statistically indistinguishable from their pre-test results does not guarantee that the post-test results would not show a significant difference when compared to native control's results. Therefore, the second procedure was to compare the native English-speaking control group's ratings and errors in all conditions to those of the non-native experimental group's post-test results. Mean ratings and percentage of errors of the native control and non-native experimental group's pre-test and post-test results in the four conditions are compared in Figure 5.25 and Figure 5.26, respectively.

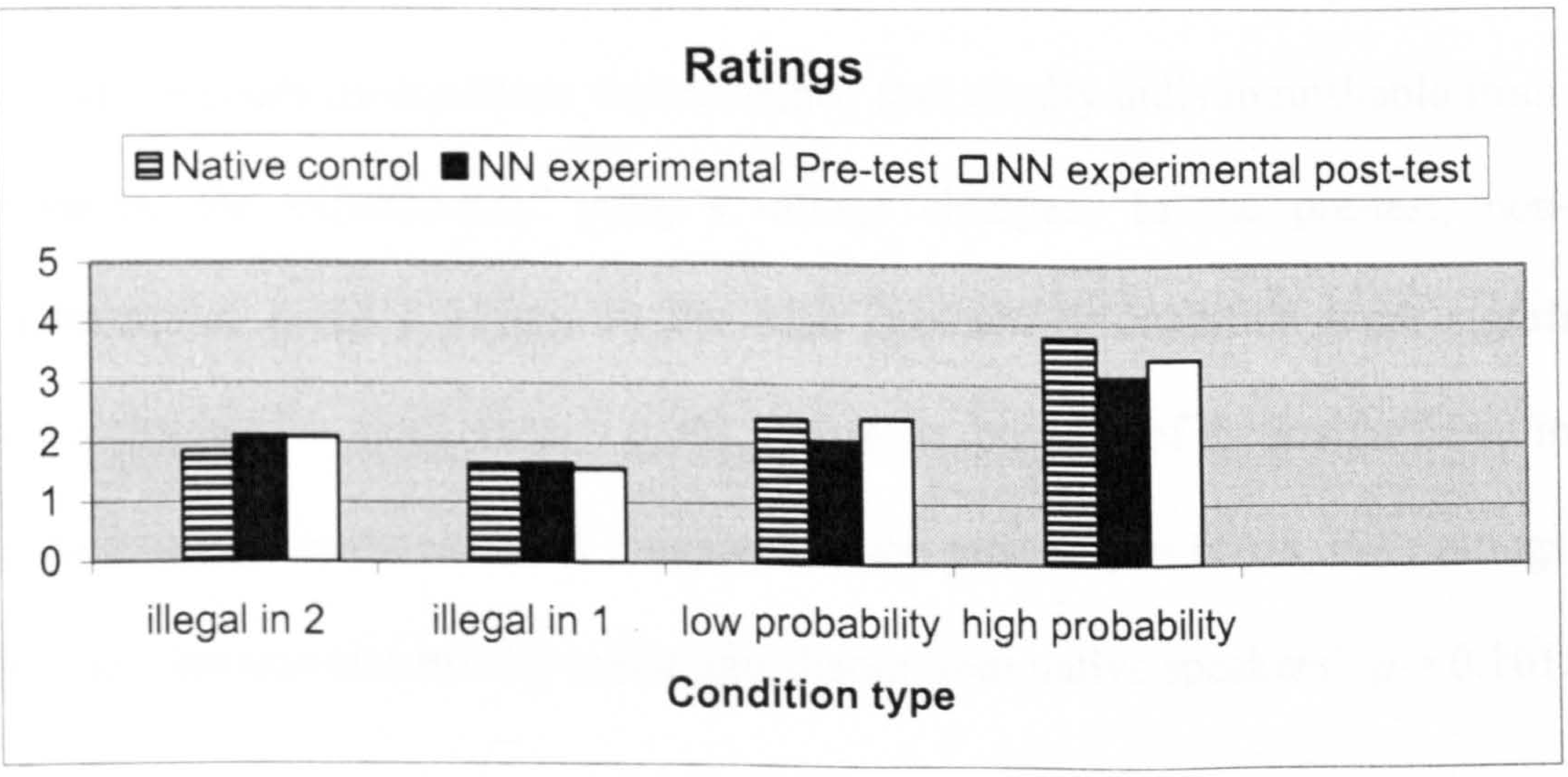

Figure 5.25 Native control and non-native experimental group's pre-test and post-test ratings in the NWR Task in four different phonotactic conditions.

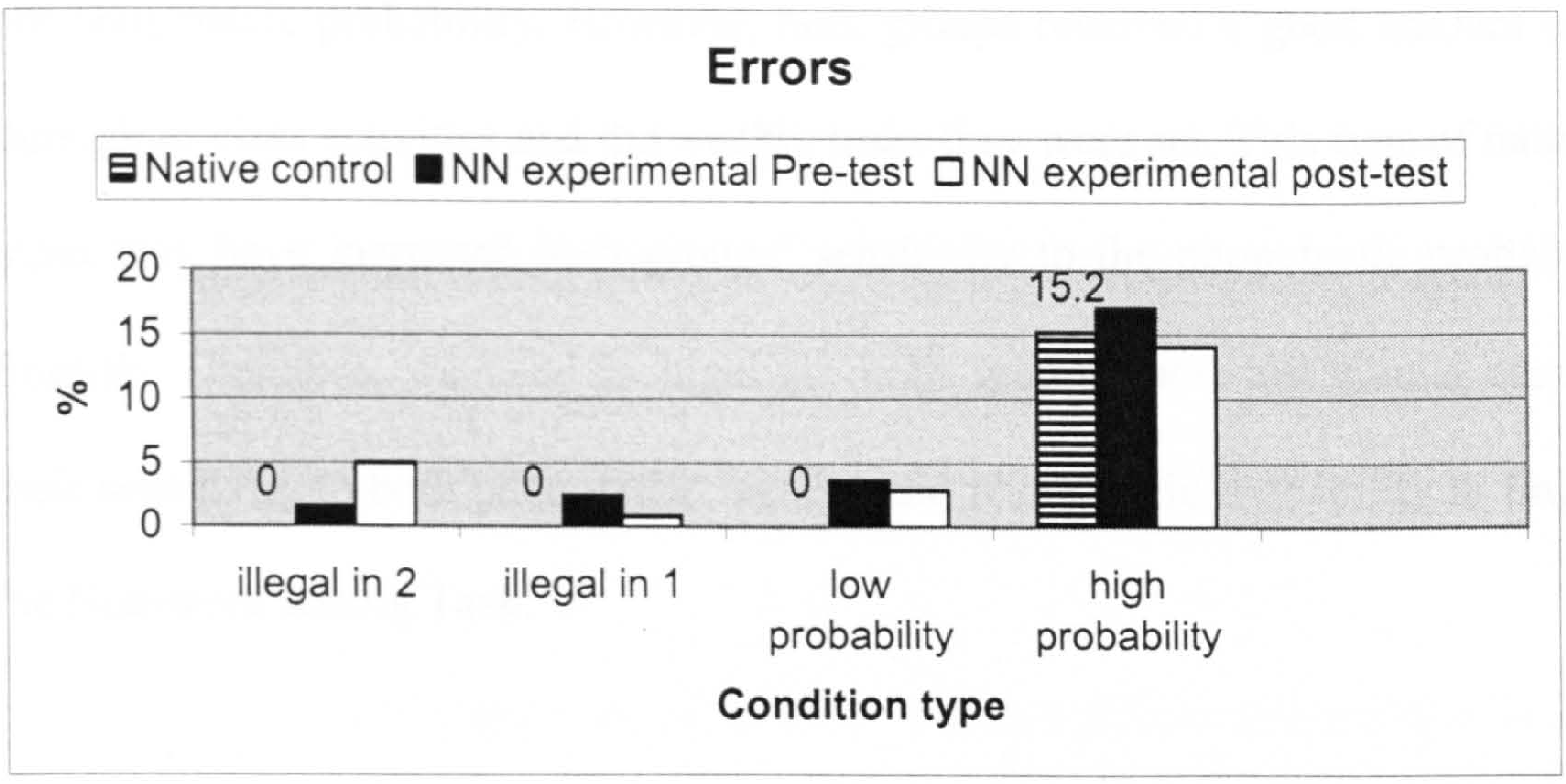

Figure 5.26 Native control and non-native experimental group's pre-test and post-test errors in the NWR Task in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the native control group's ratings (in the first ANOVA) and error rate (in the second) to those of the non-native experimental group's post-test in all four conditions. Post hoc test results showed that the non-native experimental group's ratings and error rates remained statistically indistinguishable from native speakers' in three conditions: (*illegal in 2,* Ratings, $p = 0.244$; Errors $p = 0.119$; *illegal in 1,* Ratings, $p = 0.763$; Errors $p = 0.795$; *low probability,* Ratings, $p = 0.901$; Errors $p = 0.729$). However, although the error rate in the *high probability* condition also remained statistically indistinguishable from native speakers, the experimental group's ratings changed. In the pre-test, non-native experimental group's ratings in the *high probability* condition were significantly lower than native speakers' $p = 0.001$. However, because of the insignificant increase in their ratings in the post-test compared to the pre-test $p = 0.108$, their ratings in the post-test became statistically indistinguishable from native speakers' $p = 0.101$.

5.3.1.5 Discussion

Recall that it was only the non-native experimental group who received the treatment. The treatment was aimed at teaching only the phonotactic constraints of English and not phonotactic probability. However, both groups received a good amount of input through in-class activities and the weekly tasks they were set. This type of naturalistic input may have increased both groups' sensitivity to the phonotactic probability of English. Therefore, we need to highlight both groups' post-test results concerning their sensitivity to both phonotactic legality and phonotactic probability in English in the Non-word Rating Task.

Non-native control group. Post-test results show that the non-native control group's sensitivity to English phonotactic legality did not change. Their ratings and error rates in the *illegal in 2* and *illegal in 1* conditions in the post-test remained statistically indistinguishable from theirs in the pre-test. In addition, their results did not change compared to native speakers. In the *illegal in 2* condition their ratings remained higher and their errors statistically indistinguishable from native speakers. In the *illegal in 1* conditions both their ratings and errors remained statistically indistinguishable from native speakers.

Actually, these results were not unexpected. Recall that in the pre-test the non-native control group showed sensitivity to phonotactic legality in English. They were native-like in judging non-words starting with illegal clusters in English less English- like and with fewer errors than non-words with low phonotactic probability. Therefore, there was no margin for improvement (i.e. a ceiling effect).

Post-test results also show that the non-native control group's sensitivity to English phonotactic probability did not change. Their ratings and error rates in the *low probability* and *high probability* conditions in the post-test remained statistically indistinguishable from theirs in the pre-test. In addition, their results did not change compared to native speakers. In the *low probability* conditions both their ratings and errors remained statistically indistinguishable from native speakers. Finally, in the *high probability* condition, their ratings remained significantly lower and errors statistically indistinguishable.

Again, these results were not unexpected because in the pre-test the non-native control group showed sensitivity to the phonotactic probability of English. Just like native speakers, they judged non-words with low phonotactic probability less English-like than non-words with high phonotactic probability. However, in the *high probability* condition, their ratings were significantly lower than native speakers. Therefore, the only margin for improvement which could have resulted from exposure to naturalistic input was that their ratings in the *high probability* condition may have increased to be indistinguishable from native speakers. This did not happen. However, although there was no improvement, this step is good in showing that pre-test results are valid.

Non-native experimental group. Post-test results show that the non-native control group's sensitivity to English phonotactic legality did not change. Their ratings and error rates in the *illegal in 2* and *illegal in 1* conditions in the post-test remained statistically indistinguishable from theirs in the pre-test. In addition, their results did not change compared to native speakers'. In both conditions, both their ratings and errors remained statistically indistinguishable from native speakers.

The experimental group was taught some illegal clusters in English and therefore their judgement in the Non-word Rating Task was expected to improve. However, this did not happen, although the Non-word Rating Task allows use of the sort of metalinguistic knowledge the treatment provided. This result was not unexpected after seeing the experimental group's pre-test results. In the pre-test, just like the non-native control group, the experimental group showed sensitivity to the phonotactic legality of English. They were native-like in judging non-words starting with illegal clusters in English less English-like and with fewer errors than non-words with low phonotactic probability. Therefore, there was little margin for improvement which could have resulted from explicit training.

On the other hand, Post-test results showed that there was a change in the non-native experimental group's ratings regarding their sensitivity to English phonotactic probability. Their ratings and error rates in the *high probability* condition in the post-test remained statistically indistinguishable from theirs in the pre-test. However, although their error rate in the *low probability* condition also remained statistically indistinguishable from theirs in the pre-test, their ratings became higher. In addition, their ratings and errors remained statistically indistinguishable from native speakers' in the *low probability* condition. However, because of the insignificant increase in their ratings in the *high probability* condition in the post-test, their ratings in the *high probability* condition became statistically indistinguishable from native speakers while it was significantly lower in the pre-test.

These results show that explicit teaching of illegal clusters indirectly affected the non-native experimental group's post-test ratings of non-words with low and high phonotactic probability. In other words, because they knew that non-words starting with illegal clusters should be the least English-like, they tried to differentiate between their rating of illegal non-words and those with low and high phonotactic probability. As a result, they rated non-words with low phonotactic probability and high phonotactic probability higher than in the pre-test making their ratings in the low probability condition significantly higher than theirs in the pre-test and their ratings in the high probability condition statistically indistinguishable from native speakers'.

5.3.2 Results of the Lexical Decision Task

The same analysis used in the pre-test was also used in the post-test. Practice items and real-word fillers were not included in the analysis. In this task two dependent variables were analysed, namely RTs and error rate. An error was counted when no response was made when a non-word was presented.

5.3.2.1 Comparing non-native control pre-test with post-test's results

The researcher started analyzing the non-native control group's post-test results of the Lexical Decision Task by comparing them to pre-test results in all four conditions. The non-native control group's mean RTs and mean percentage of errors in four different phonotactic conditions in the pre-test and post-test are compared in Figure 5.27 and Figure 5.28, respectively.

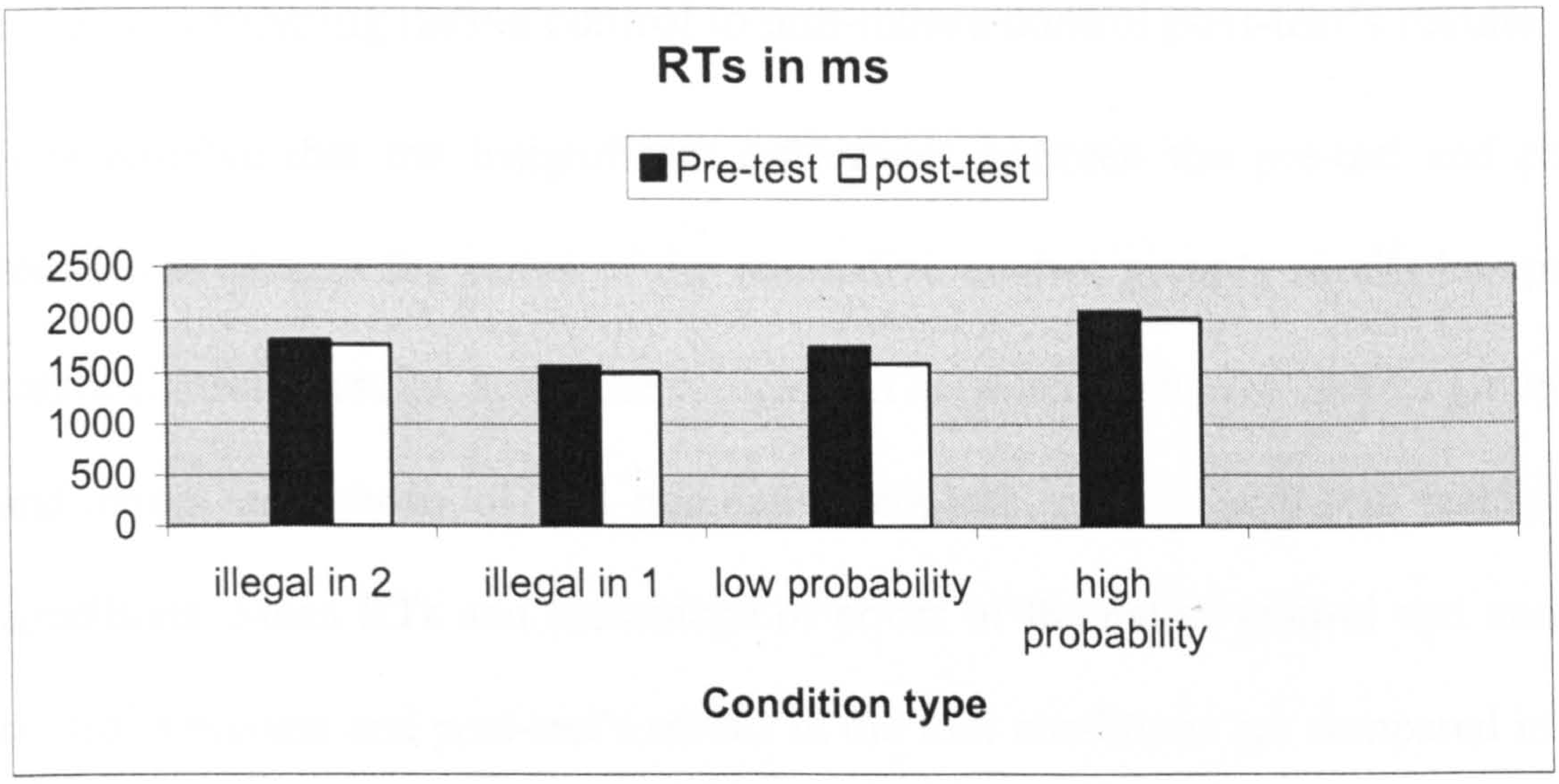

Figure 5.27 Non-native control group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

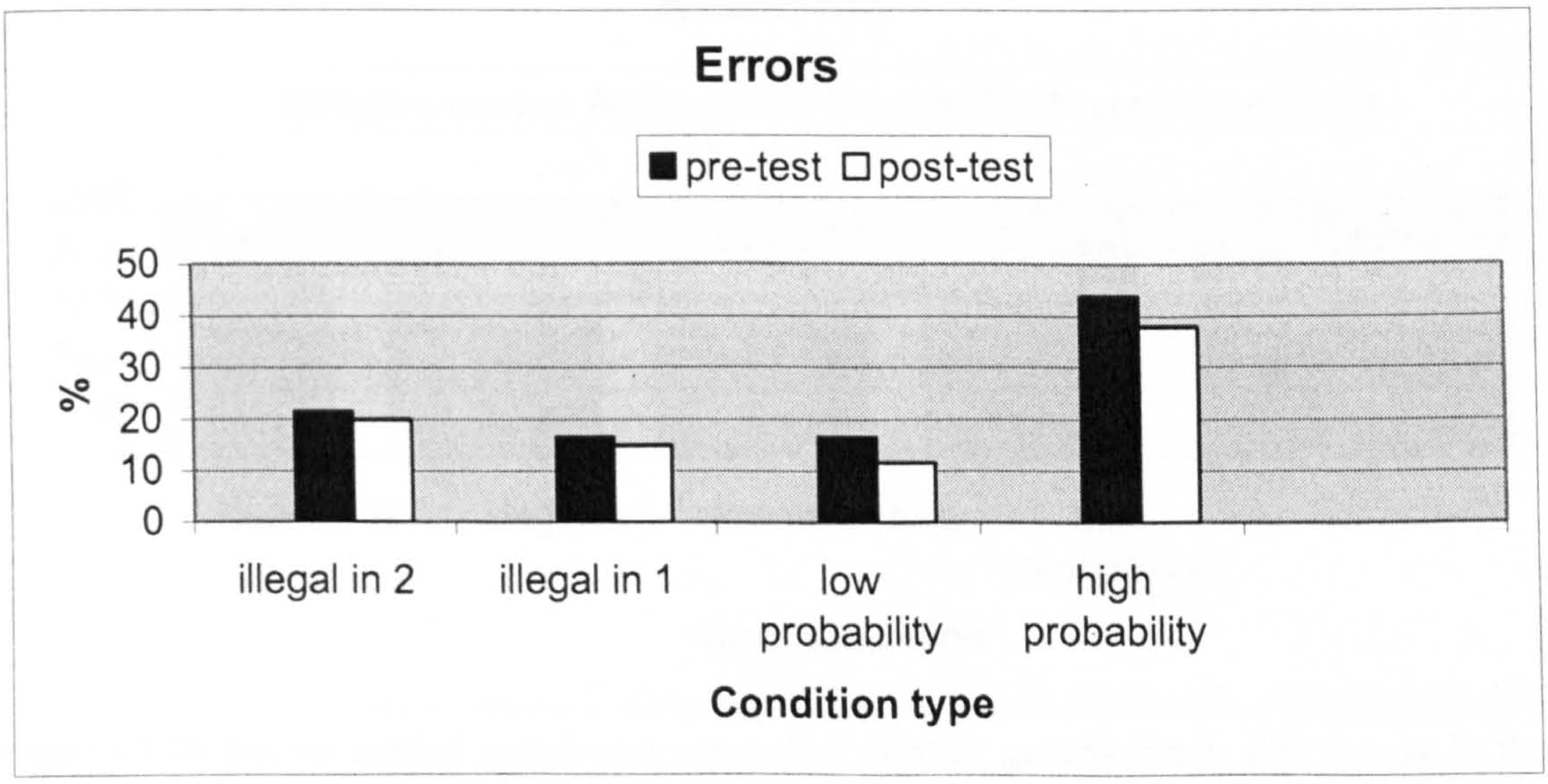

Figure 5.28 Non-native control group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the non-native control group's post-test RTs (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed the non-native control group's RTs and errors in the post-test in all conditions remained statistically indistinguishable from theirs in the post-test (*illegal in 2* condition (RTs, $p = 0.722$; Errors $p = 0.725$), *illegal in 1* condition (RTs, $p = 0.589$; Errors $p = 0.725$), *the low probability* condition (RTs, $p = 0.183$; Errors $p = 0.291$) and *the high probability* condition (RTs, $p = 0.568$; Errors $p = 0.219$).

5.3.2.2 Comparing native control to non-native control post-test's results

It is possible that the insignificant differences between the pre-test and post-test results can change the status of the non-native control group's results compared to native control's results, hence the comparison between the native control group's RTs and errors and those of the non-native control group's post-test results in all conditions. Mean RTs and percentage of errors of the native control and non-native control's pre-test and post-test's results in the four conditions are compared in Figure 5.29 and Figure 5.30, respectively.

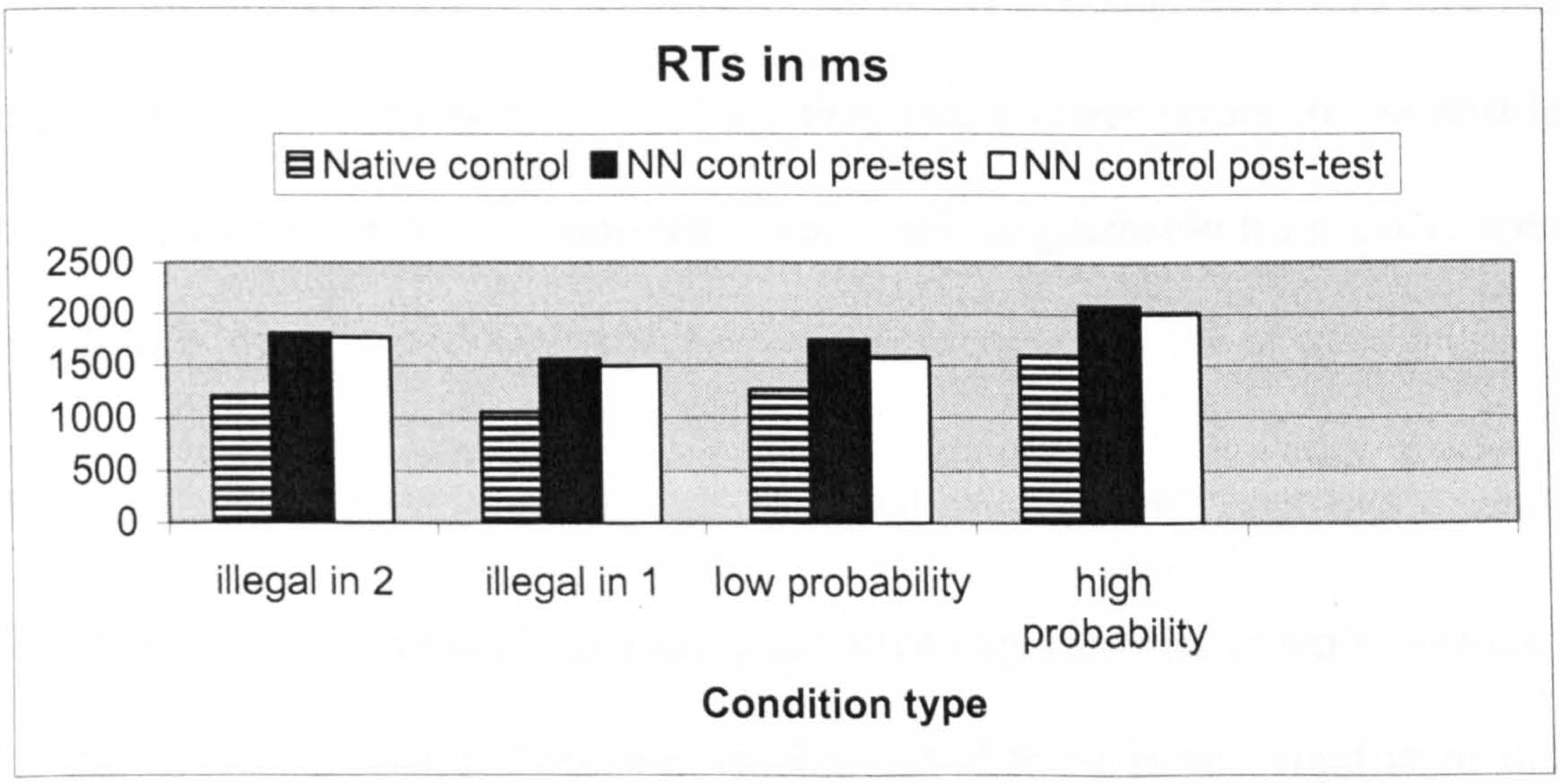

Figure 5.29 Native control group and non-native control group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

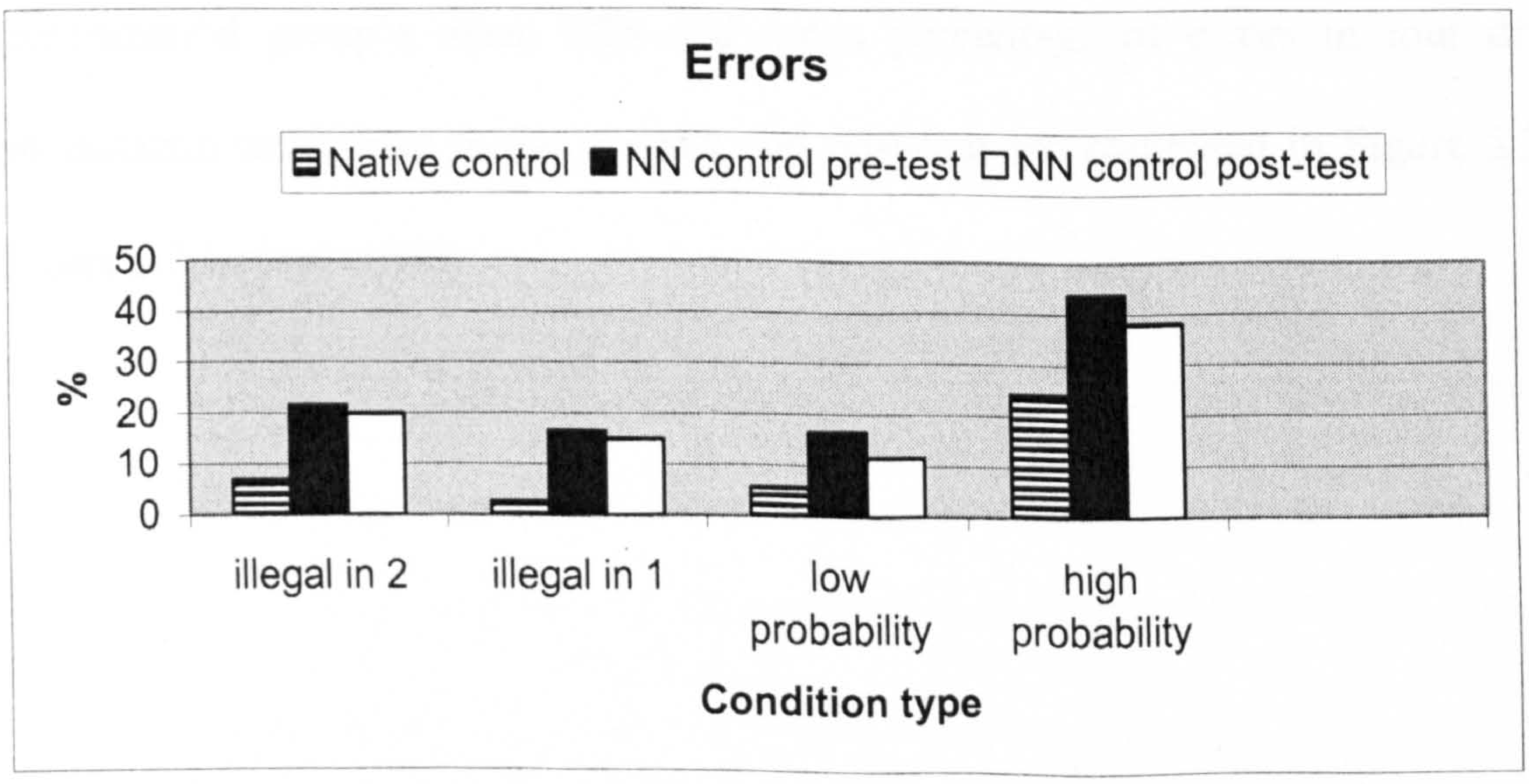

Figure 5.30 Native control group and non-native control group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the native control group's RTs (in the first ANOVA) and error rate (in the second) to those of the non-native control group's post-test's in all four conditions. Post hoc test results showed that the non-native control group's results did not change in the post-tests compared to native speakers' in three conditions. In the *illegal in 2, illegal in 1* and *high probability* conditions their RTs and errors remained higher than native speakers' (*illegal in 2* condition (RTs, $p = 0.000$; Errors $p = 0.018$), *illegal in 1* condition (RTs, $p = 0.002$; Errors $p = 0.026$) and *the high probability* condition (RTs, $p = 0.004$; Errors $p = 0.013$). However, in the *low probability* condition although their RTs also remained higher than native speakers $p = 0.031$, they made fewer errors in the post-test and therefore their error rate became statistically indistinguishable from native speakers $p = 0.264$.

5.3.2.3 Comparing non-native experimental pre-test with post-test's results

The first procedure in analyzing the non-native experimental group's post-test results of the Lexical Decision Task was to find out if there is any significant difference between pre-test and post-test results in all four conditions. The non-native experimental group's mean RTs and mean percentage of errors in four different phonotactic conditions in the pre-test and post-test are compared in Figure 5.31 and Figure 5.32, respectively.

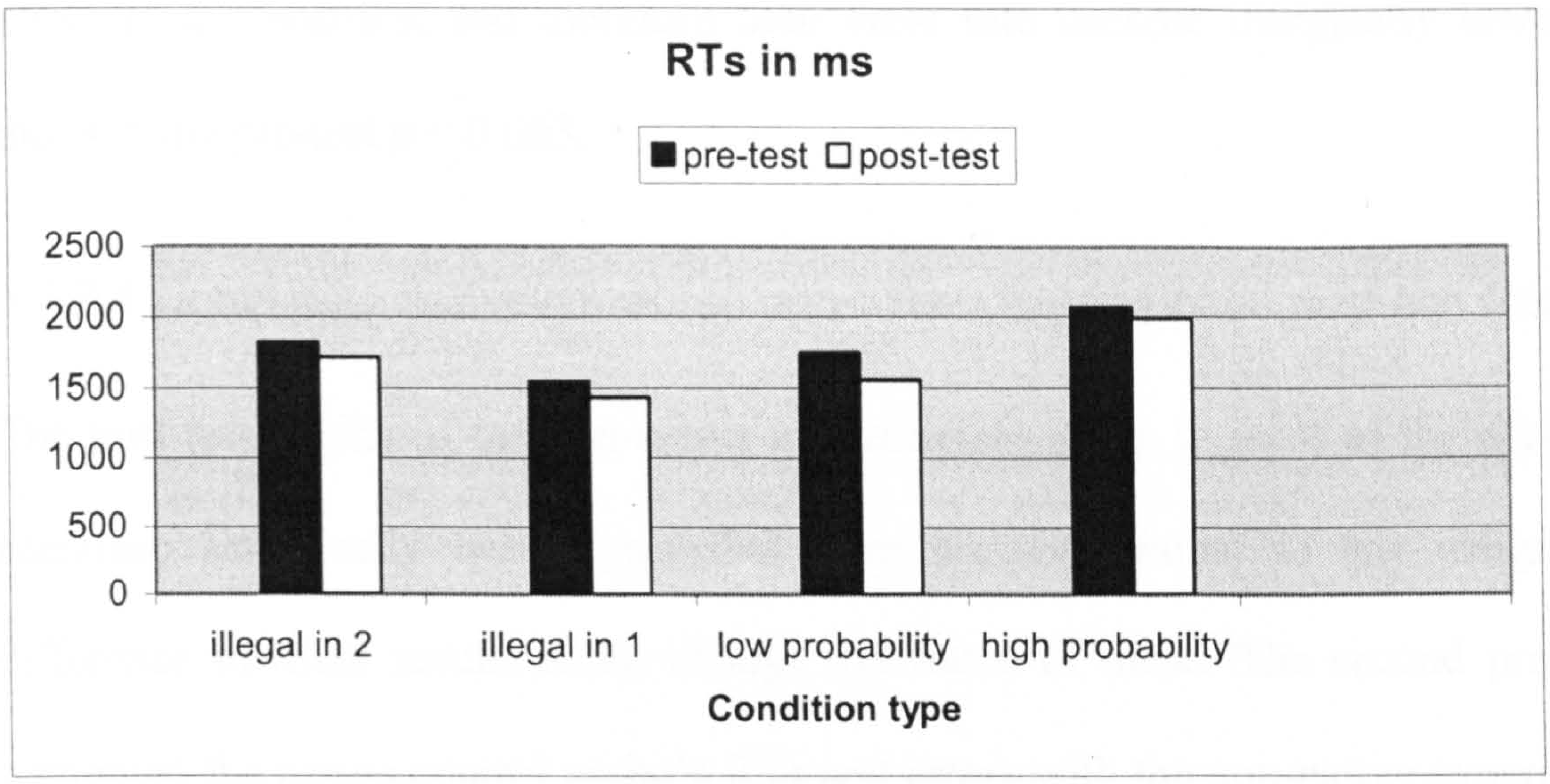

Figure 5.31 Non-native experimental group's mean RTs in ms in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

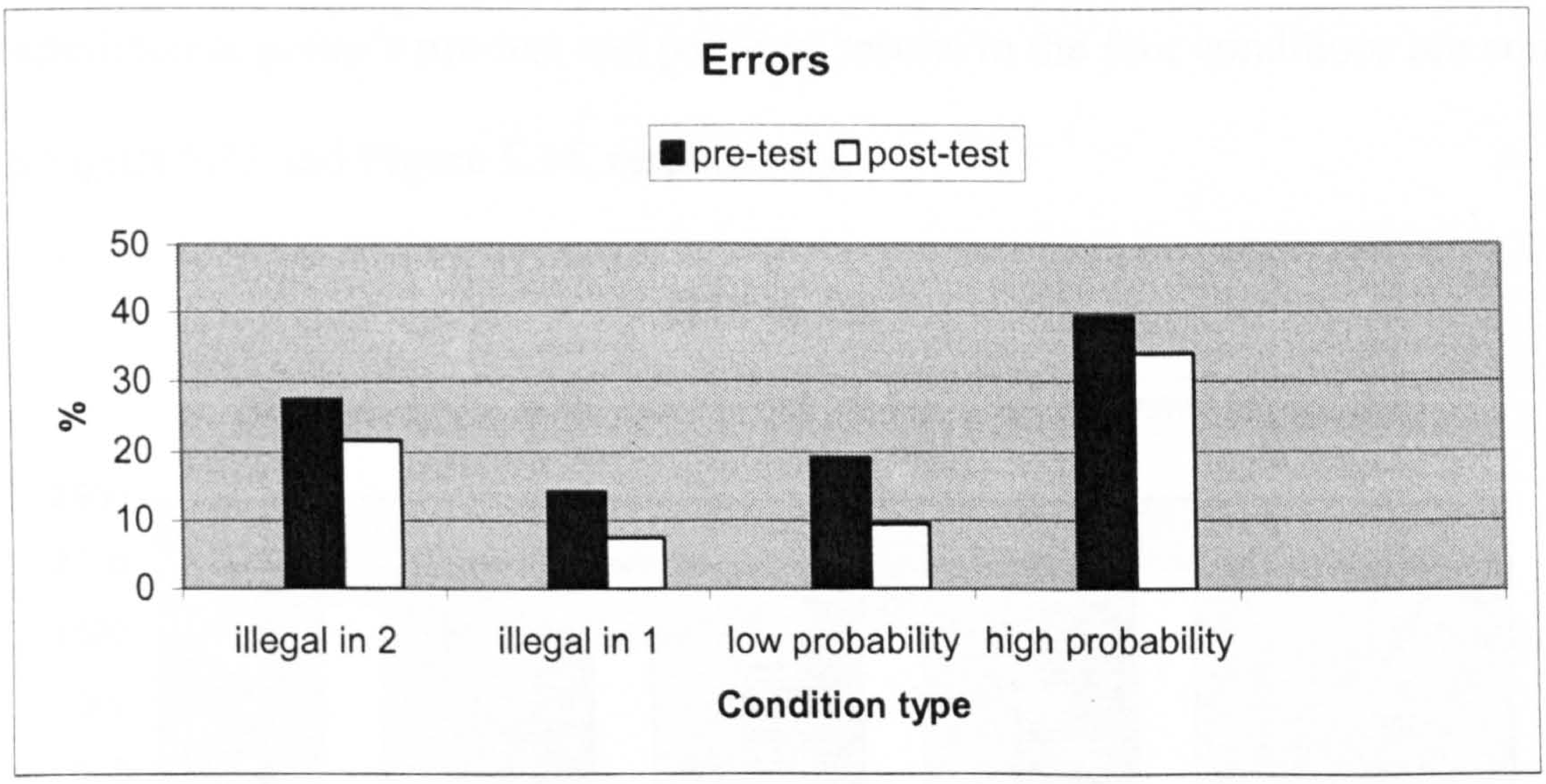

Figure 5.32 Non-native experimental group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the non-native experimental group's post-test's RTs (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed that the non-native experimental group's RTs and errors in the post-test in three conditions remained statistically indistinguishable from theirs in the post-test (*illegal in 2* condition (RTs, $p = 0.407$; Errors $p = 0.256$), *illegal in 1* condition (RTs, $p = 0.377$; Errors $p = 0.194$), and *the high probability* condition (RTs, $p = 0.550$; Errors $p = 0.291$). However, in the *low probability* condition, although their RTs also remained statistically indistinguishable from theirs in the pre-test $p = 0.126$, they made fewer errors in the post-test and therefore their error rate became marginally lower than theirs in the pre-test $p = 0.063$.

5.3.2.4 Comparing native control to non-native experimental post-test results

The post-test results of the non-native experimental group in most of the conditions remained statistically indistinguishable from pre-test results, so any insignificant difference in these results could change the status of these. The second procedure compared the native control group's RTs and errors with the non-native experimental group's. Mean RTs and percentage of errors of the native control and non-native experimental group's pre-test and post-test results in the four conditions are compared in Figure 5.33 and Figure 5.34, respectively.

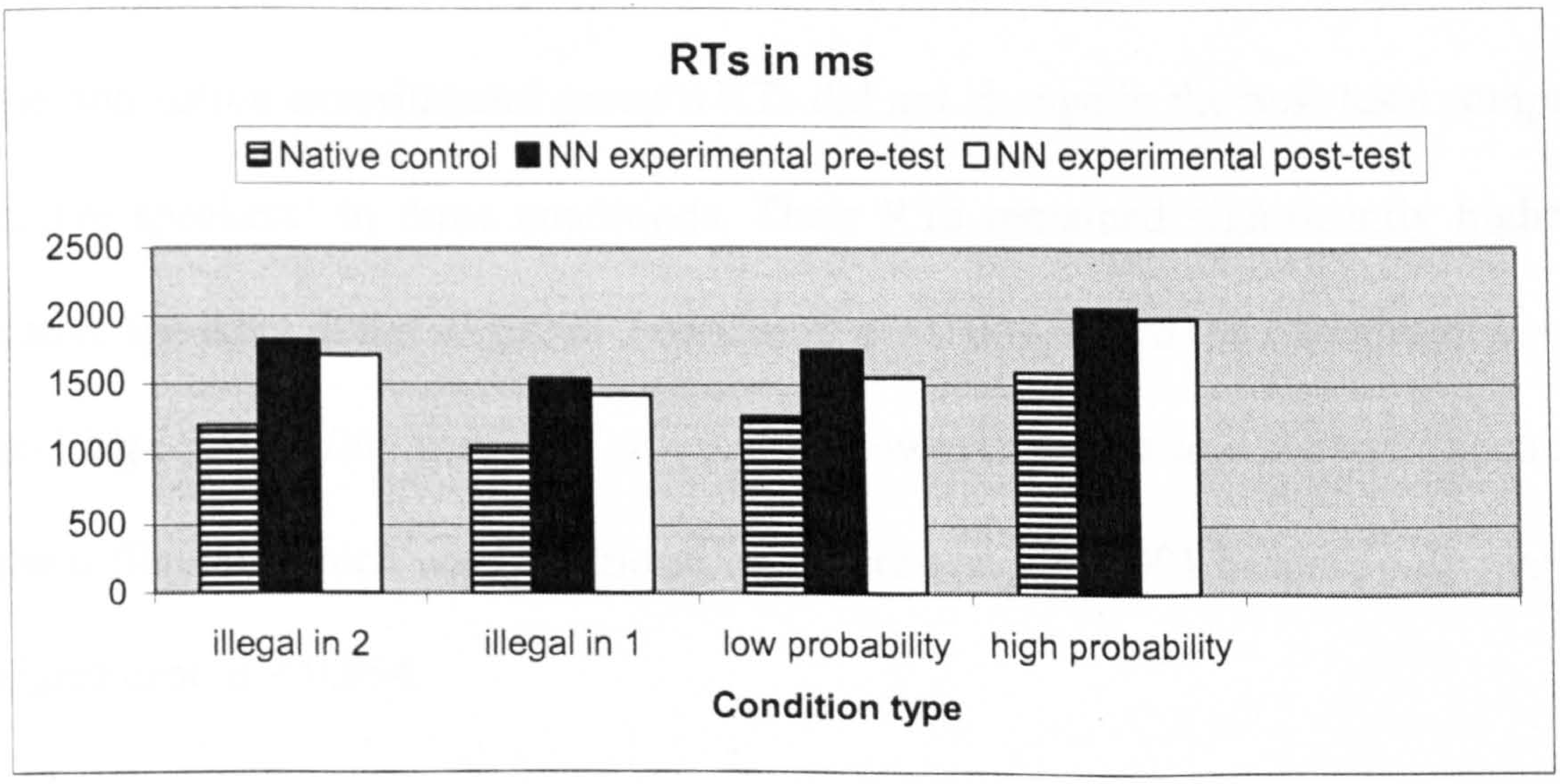

Figure 5.33 Native control group and non-native experimental group's mean RTs in ms in the pre-test and post-test in four different phonotactic conditions.

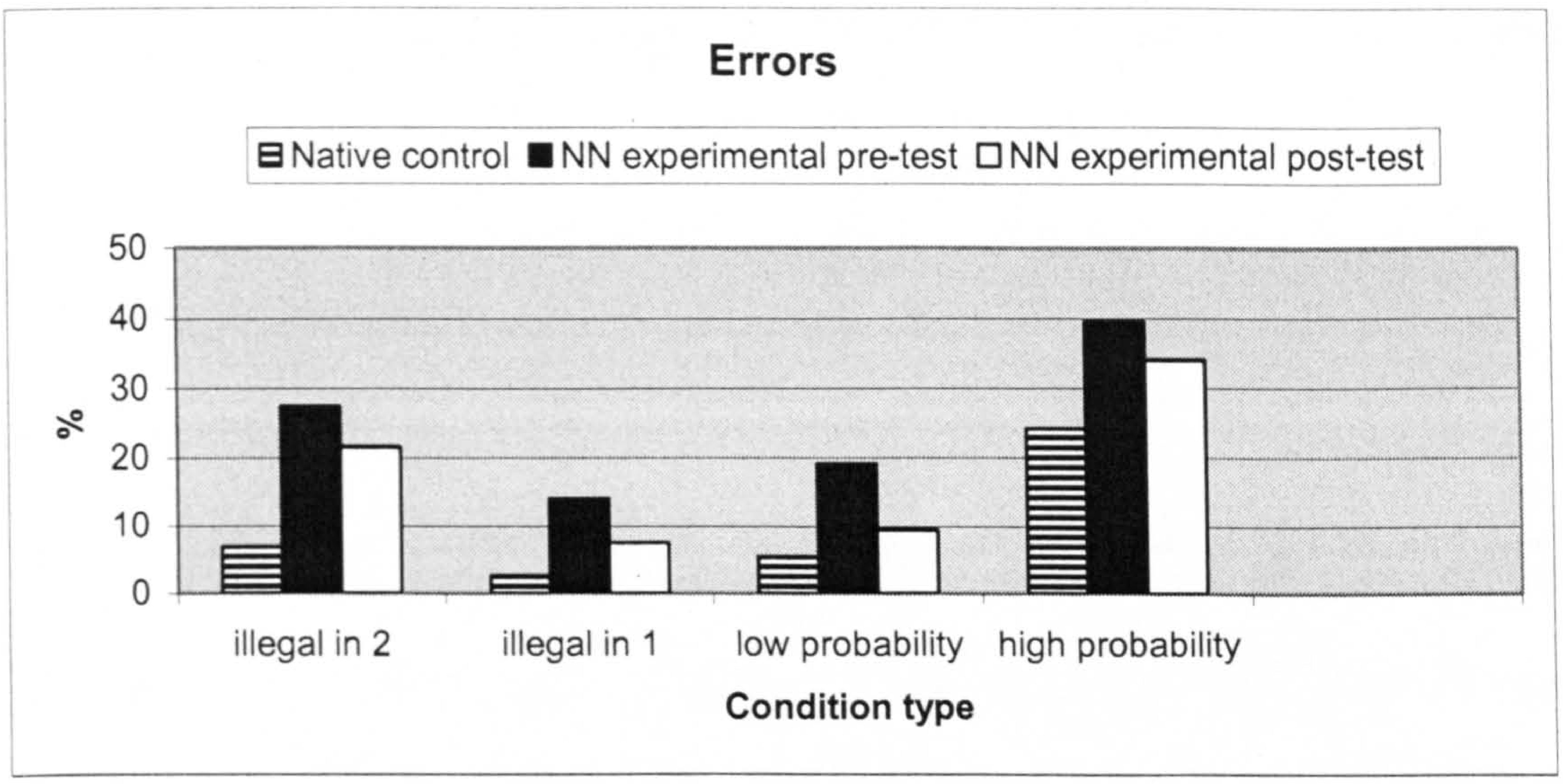

Figure 5.34 Native control and non-native experimental group's mean percentage of errors in the Lexical Decision Task in the pre-test and post-test in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the native control group's RTs (in the first ANOVA) and error rate (in the second) to those of the non-native experimental group's post-test in all four conditions. Post hoc test results showed that the non-native experimental group's RTs did not change in the post-tests compared to native speakers' in three conditions. Their RTs remained significantly higher than native speakers in the *illegal in 2* condition $p = 0.001$, *illegal in 1* condition $p = 0.012$, and *high probability* condition $p = 0.008$. However, in the *low probability* condition, the difference which was significant in the pre-test $p = 0.001$ became only marginally significant $p = 0.054$.

On the other hand, there were some differences as far as the error rate is concerned. Their error rate in the *illegal in 2* condition remained significantly higher than native speakers $p = 0.014$. However, in the *illegal in 1* condition, *low probability* condition, and *high probability* condition, their error rates which were marginally higher in the *illegal in 1* $p = 0.064$ condition and significantly higher in the *low probability* condition $p = 0.027$ and *high probability* condition $p = 0.013$ became statistically indistinguishable from the *illegal in 1* condition $p = 0.425$ and *low probability* condition $p = 0.496$ and only marginally higher in the *high probability* condition $p = 0.097$.

5.3.2.5 Discussion

Although the treatment which the non-native experimental group received did not target phonotactic probability, the amount of oral naturalistic input both groups received through the in-class activities in English and the weekly tasks they were set may have increased both groups' sensitivity to phonotactic probability in English. Therefore, we need to highlight both groups' post-test results concerning their sensitivity to both phonotactic legality and phonotactic probability of English in the Lexical Decision Task.

Two important points first have to be noted here. First, recall that in the pre-test, the Lexical Decision Task failed to show that non-native speakers were sensitive to the phonotactic legality of either English or Arabic. In the pre-test, our native speakers and non-native speakers in both groups rejected non-word stimuli items in the *low probability* condition as fast and with statistically indistinguishable number of errors as they did with those in the *illegal in 2* condition and *illegal in 1* condition. One of the explanations I provided was that in the auditory Lexical Decision Task non-words can not be rejected earlier based on the initial illegal cluster and that an evaluation of the whole item instead seems to take place. Therefore, in the post-test the Lexical Decision Task is not designed to show that there is an improvement of subjects' sensitivity to phonotactic legality. However, it will be good to see that pre-test results are duplicated.

The second point concerns development of subjects' sensitivity to phonotactic probability. If subjects' sensitivity to the phonotactic probability develops in the post-test we would expect longer RTs and more errors in the *high probability* condition. However, a second presentation in the post-test, although separated by at least eight weeks, may have actually familiarised subjects with the items and therefore produced a counter effect (i.e. shorter RTs and less errors). This should be born in mind. I now turn to both non-native groups' post-test results.

Non-native control group. Post-test results show that the non-native control group's sensitivity to English phonotactic legality did not change. Their RTs and error rates in the *illegal in 2* and *illegal in 1* conditions in the post-test remained statistically indistinguishable from theirs in the pre-test. In addition, their results did not change compared to native speakers. As in the pre-test, in both conditions they rejected non-words more slowly and with more errors than native speakers.

On the other hand, post-test results show that there was a slight change in the non-native control group's sensitivity to English phonotactic probability. Their RTs and error rates in the *low probability* and *high probability* conditions in the post-test remained statistically indistinguishable from theirs in the pre-test. However, because of the insignificant decrease of their errors in the *low probability* condition, their results changed when compared to native speakers. Although they still rejected non-words with low phonotactic probability more slowly and those with high phonotactic probability both more slowly and with more errors than native speakers, their errors in the *low probability* condition became statistically indistinguishable from native speakers'. In a way this might be considered as a development of sensitivity to phonotactic probability, i.e. accurately rejecting non-words with low phonotactic probability more often than in the pre-test.

Had the decrease been a mere test effect (i.e. benefiting from being familiar with the test at post-test) , we would expect similar reduction of errors in the *high probability* condition. However, a closer look at the percentage of reductions in both conditions shows that reductions in the *low probability* condition are much greater than those in the *high probability* condition (high probability condition reductions in RTs = 3.4% and in errors = 15% vs. low probability condition reductions in RTs = 10% and in errors = 43%). These results show that, whereas there seems to be a test effect that would be counter to the direction of development of the sensitivity to phonotactic probability, this effect could not prevent this development of sensitivity from showing as greater reductions in RTs and errors in the low phonotactic probability than those in the high phonotactic probability.

Non-native experimental group. Post-test results show that the non-native experimental group's RTs and error rates in the *illegal in 2* and *illegal in 1* conditions did not change when compared to their pre-test results. However, because of the insignificant decrease of errors in the post-test in the *illegal in 1* condition their error rate became statistically indistinguishable from native speakers. This actually shows that the non-native experimental group benefited from explicit teaching of English illegal clusters and therefore more often rejected non-words starting with illegal consonant clusters, although not significantly, than in the pre-test, rendering their error rate in this condition indistinguishable from native speakers. Recall that the illegal clusters included in the list given to experimental subjects before the treatment were only those used to form the *illegal in 1* condition. Therefore, any improvement was expected to be in this condition only and not in the *illegal in 2* condition where clusters used were not given to subjects. Also, note that an improvement in RTs was not expected, as I argued above that during an auditory lexical decision, a decision seems to be given based on an evaluation of the whole item and not only the initial illegal cluster.

Post-test results show that the non-native experimental group's sensitivity to English phonotactic probability also changed. First, they made marginally fewer errors in the *low probability* condition than in the pre-test. This lower error rate made their errors in the post-test statistically indistinguishable from native speakers. In addition, their RTs in the *low probability* condition, which were significantly slower than native speakers', became only marginally slower. These results allow us to partially reject the null Hypothesis 14, which states that "Control and experimental non-native subjects will show no improvement towards their sensitivity to the phonotactic probability and phonotactic legality of English at post-test". These results show that the experimental subjects' sensitivity to English phonotactic probability has developed by the post-test. They rejected non-words with low phonotactic probability marginally more often and faster (though not significantly) than in the pre-test. This made their errors in the *low probability* condition statistically indistinguishable from native speakers' and their RTs only marginally slower than native speakers'. Again, this is not a mere test effect as we would expect similar reductions in RTs and errors in the high probability condition. As in the non-native control group's results, a closer look at the percentage of reductions in RTs and error rates in both conditions shows that reductions in the low probability condition are much greater than those in the *high probability* condition (*high probability* condition reductions in RTs = 3.8% and in errors = 14.7% vs. *low probability* condition reductions in RTs = 12.5% and in errors = 50%). These results duplicate results from the non-native control group which show that, whereas there seems to be a test effect that is counter to the direction of development of the sensitivity to phonotactic probability, this effect could not prevent this development of sensitivity from showing as greater reductions in RTs and errors in low phonotactic probability than those in high phonotactic probability.

5.3.3 Results of the Word Spotting Task

The same coding and analysis procedures used in the pre-test were also used in the post-test. RTs and error rates were counted and analysed. Practice items and real-word fillers were excluded from analysis.

5.3.3.1 Comparing non-native control pre-test with post-test's results

The first procedure compared pre-test and post-test results in all four boundary conditions. The non-native control group's mean RTs and mean percentage of errors in four different boundary conditions in the pre-test and post-test are compared in Figure 5.35 and Figure 5.36, respectively.

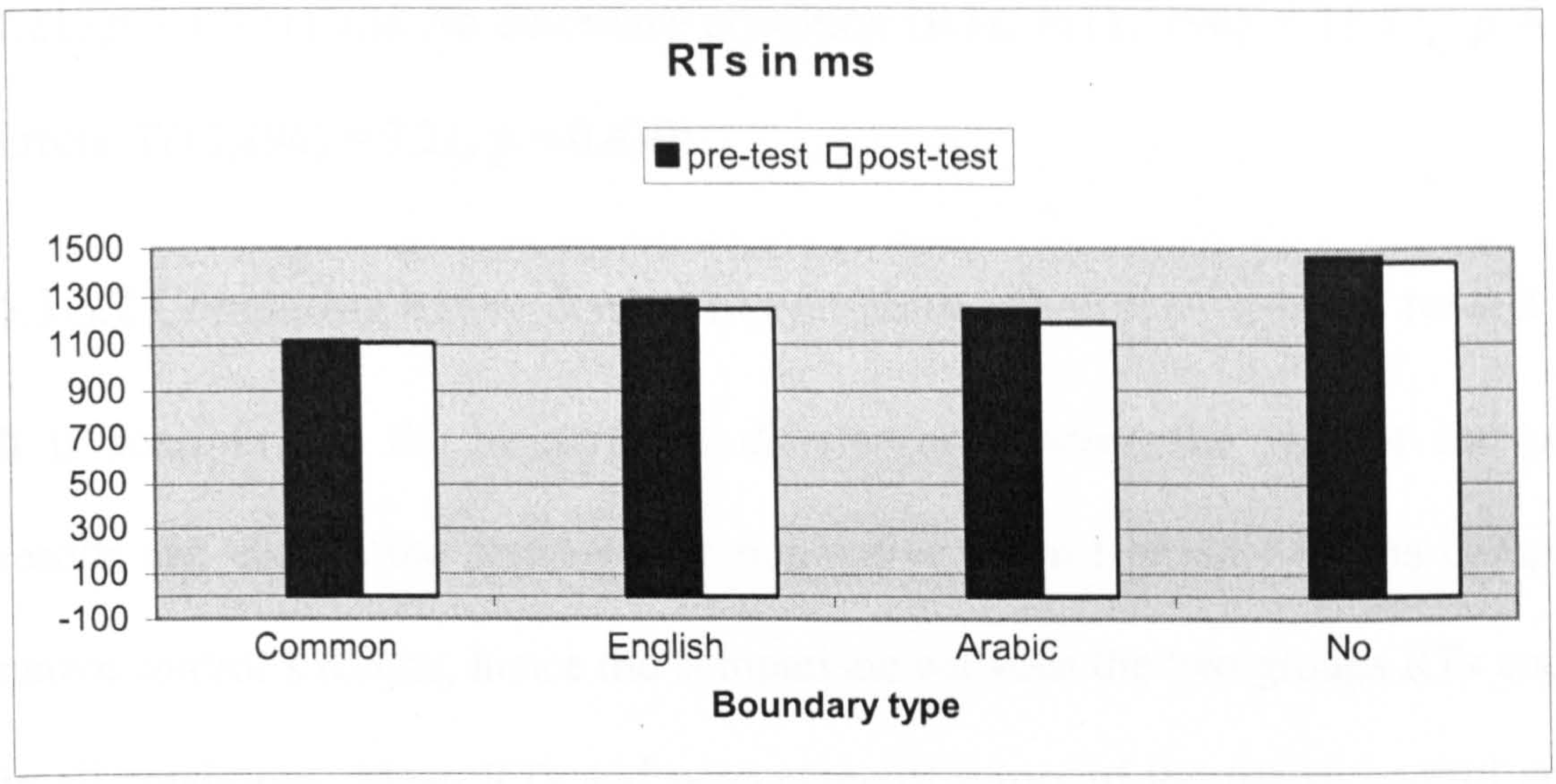

Figure 5.35 Non-native control group's mean RTs in ms in the Word Spotting Task in the pre-test and post-test in four different boundary conditions.

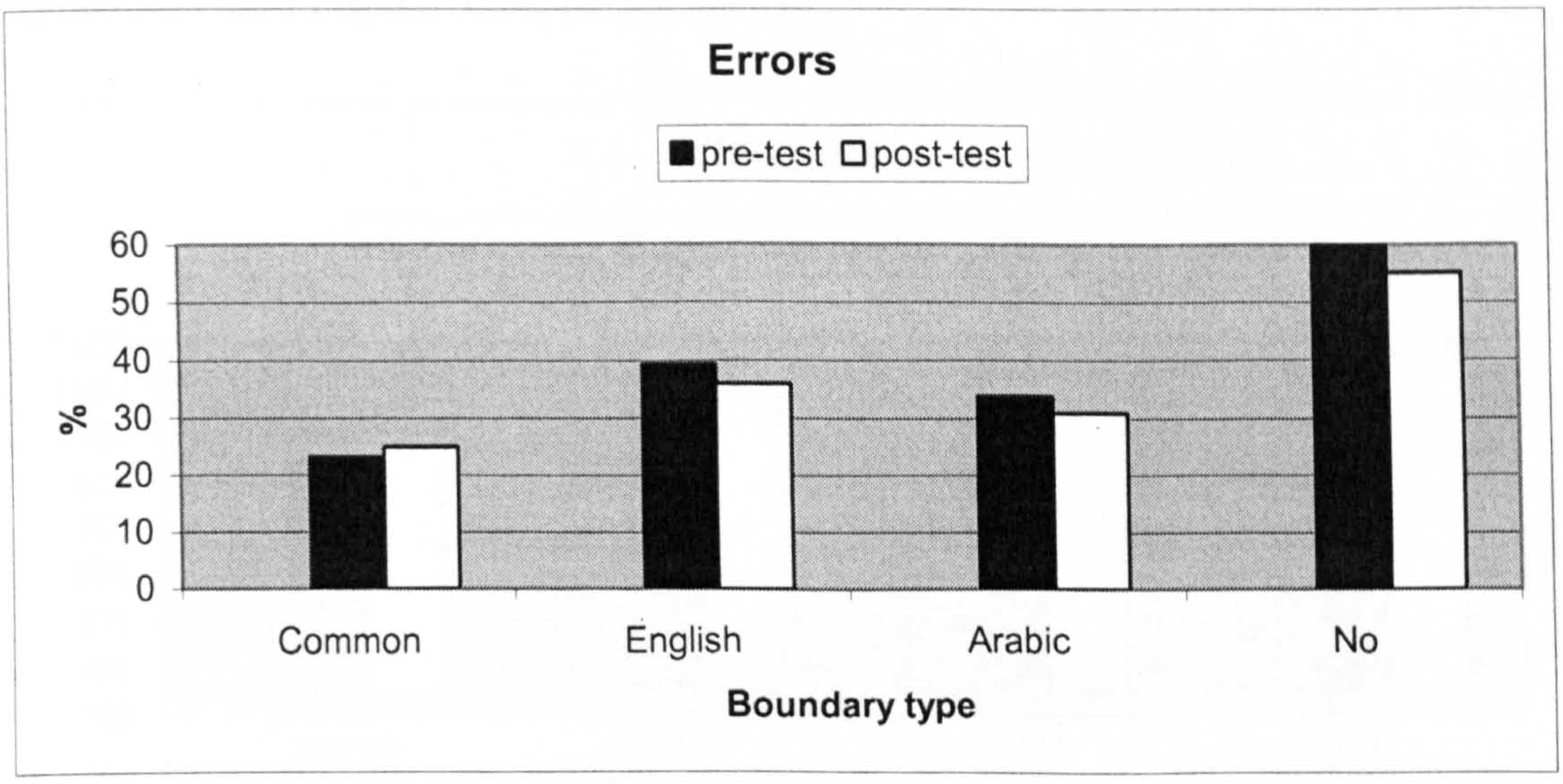

Figure 5.36 Non-native control group's mean percentage of errors in the Word Spotting Task in the pre-test and post-test in four different boundary conditions.

Two one-way ANOVAs were conducted to compare the non-native control group's post-test RTs (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed the non-native control group's RTs and errors in the post-test in all conditions remained statistically indistinguishable from theirs in the pre-test (*common boundary* condition (RTs, F(11, 196) = 11.87, p = 0.920; Errors F(11,196) = 7.21, p = 0.796), *English boundary* condition (RTs, F(11, 196) = 11.87, p = 0.715; Errors F(11,196) = 7.21, p = 0.667), *Arabic boundary* condition (RTs, F(11, 196) = 11.87, p = 0.537; Errors, F(11,196) =

7.21, $p = 0.731$) and *No boundary* condition (RTs, F(11, 196) = 11.87, $p = 0.768$; Errors, F(11,196) = 7.21, $p = 0.439$).

5.3.3.2 Comparing native control to non-native control post-test's results

It is possible that the insignificant differences between the pre-test and post-test results can change the status of the non-native control group's results compared to native control's results, hence the comparison between the two groups RTs and errors in all conditions. Mean RTs and percentage of errors of the native control and non-native control's pre-test and post-test results in the four conditions are compared in Figure 5.37 and Figure 5.38, respectively.

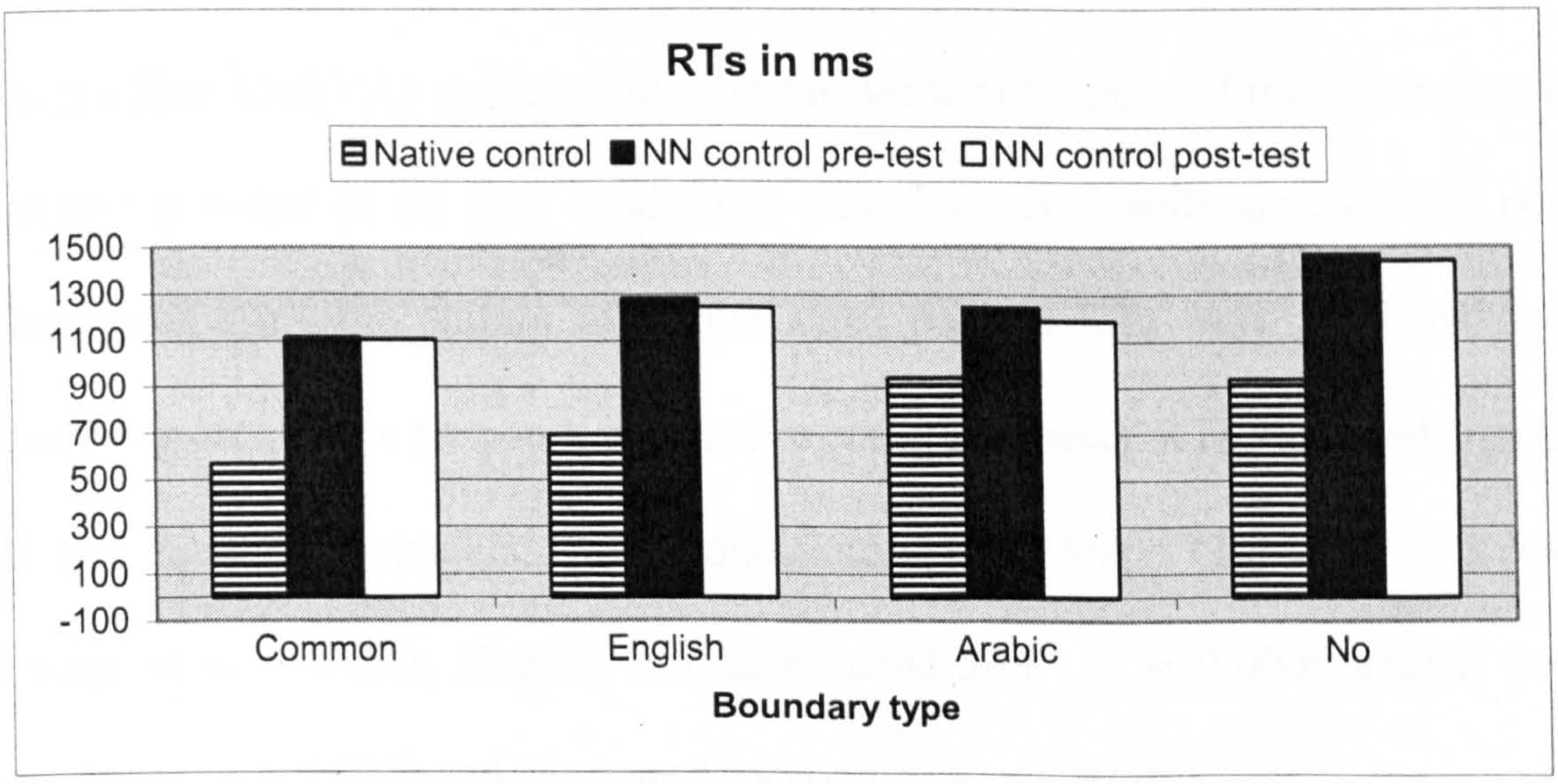

Figure 5.37 Native control group and non-native control group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different boundary conditions.

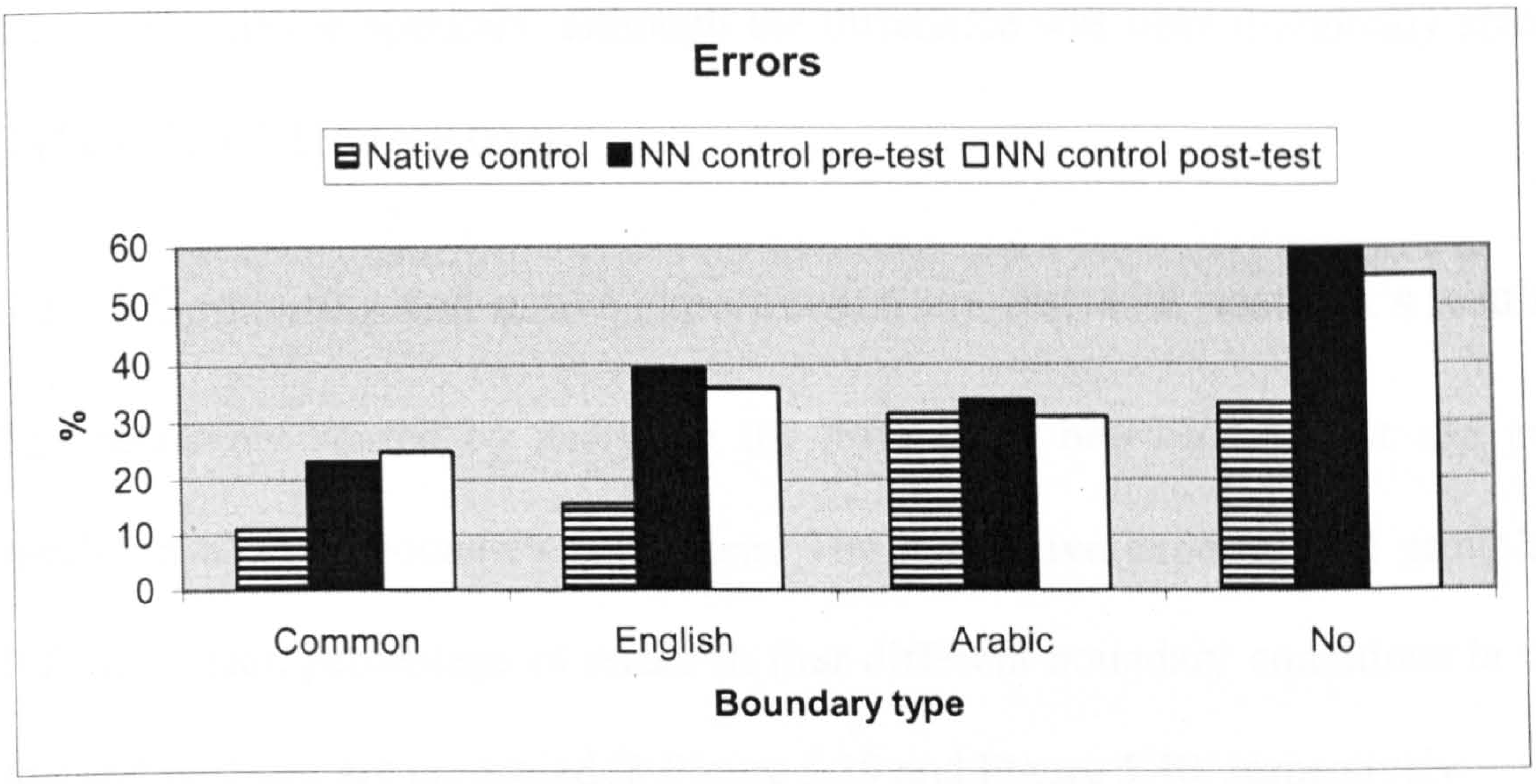

Figure 5.38 Native control group and non-native control group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different boundary conditions.

Two one-way ANOVAs were conducted to compare the native control group's RTs (in the first ANOVA) and error rate (in the second) to those of the non-native control group's post-test in all four conditions. Post hoc test results showed that post-tests duplicated the same pattern of results when pre-test RTs were compared to native control group. In the post-test, non-native control subjects' RTs remained significantly slower than native speakers' in all conditions (F(11, 196) = 11.87, *common boundary* condition p = 0.000; *English boundary* condition, p = 0.000; *Arabic boundary* condition p = 0.028; and *No boundary* condition p = 0.000).

In addition, their error rate remained significantly higher (F(11,196) = 7.21) in the *English boundary* condition p = 0.005 and *No boundary* condition p = 0.001 than native speakers' error rate in the same conditions. It also remained statistically indistinguishable in the *Arabic boundary* condition from native speakers' error rate in the same condition p = 0.980. However, because of the increase in errors in the post-test in the *Common boundary* condition the non-native control group's errors became higher than native speakers' although the difference was only marginally significant (F(11,196) = 7.21, p = 0.064).

5.3.3.3 Comparing non-native experimental pre-test with post-test's results

The researcher started by analysing the differences between pre-test and post-test results in all four boundary conditions. The non-native experimental group's mean RTs and mean percentage of errors in four different boundary conditions in the pre-test and post-test are compared in Figure 5.39 and Figure 5.40, respectively.

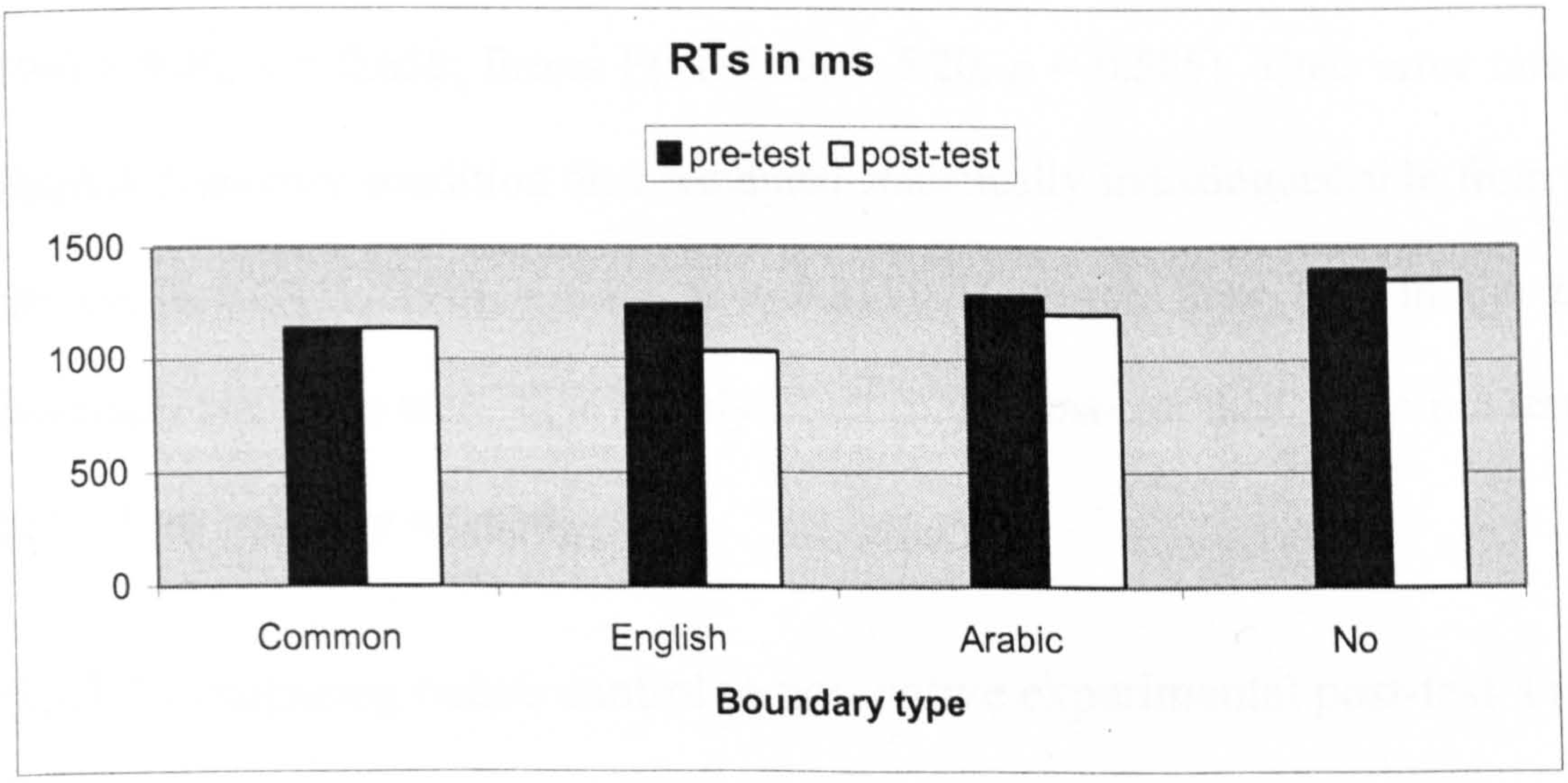

Figure 5.39 Non-native experimental group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different phonotactic conditions

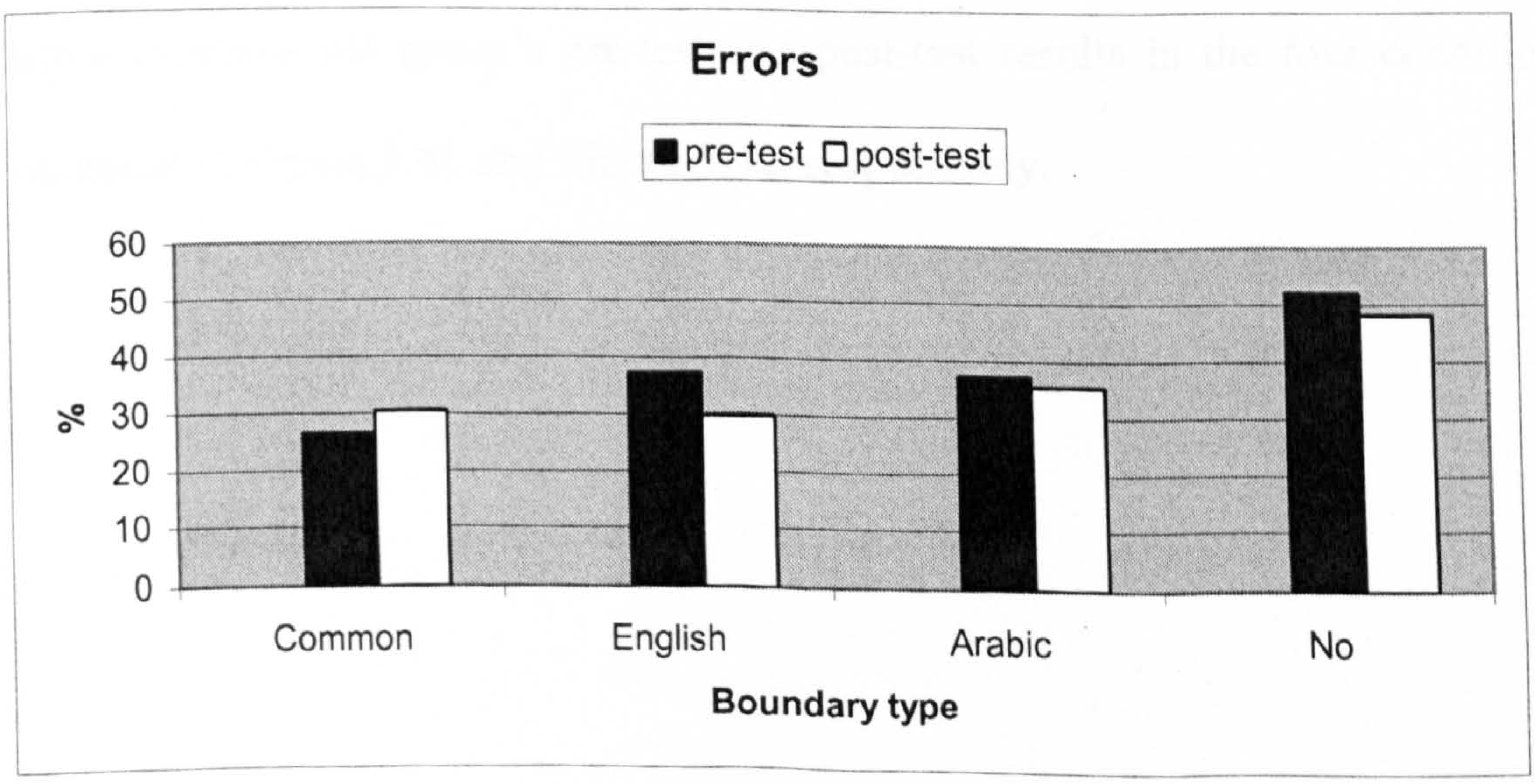

Figure 5.40 Non-native experimental group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different phonotactic conditions.

Two one-way ANOVAs were conducted to compare the non-native experimental group's post-test RTs (in the first ANOVA) and error rate (in the second) to those of their pre-test in all four conditions. Post hoc test results showed the non-native experimental group's RTs and errors in the post-test remained statistically indistinguishable from theirs in the pre-test in three conditions. These conditions include *Common boundary* condition (RTs, F(11, 196) = 9.40, p = 0.993; Errors, F(11, 196) = 5.20, p = 0.585), *Arabic boundary* condition (RTs, F(11, 196) = 9.40, p = 0.440; Errors F(11, 196) = 5.20, p = 0.815) and *No boundary* condition (RTs, F(11, 196) = 9.40, p = 0.658; Errors F(11, 196) = 5.20, p = 0.585). Their error rate in the *English boundary* condition also remained statistically indistinguishable from that in the pre-test (F(11, 196) = 5.20, p = 0.311). However, their RTs in the *English boundary* condition were significantly faster in the post-test than in the pre-test (RTs, F(11, 196) = 9.40, p = 0.034).

5.3.3.4 Comparing native control to non-native experimental post-test's results

Here the analysis was to find out how the native group compare to the experimental group at post-test. Mean RTs and percentage of errors of the native control and non-native experimental group's pre-test and post-test results in the four conditions are compared in Figure 5.41 and Figure 5.42, respectively.

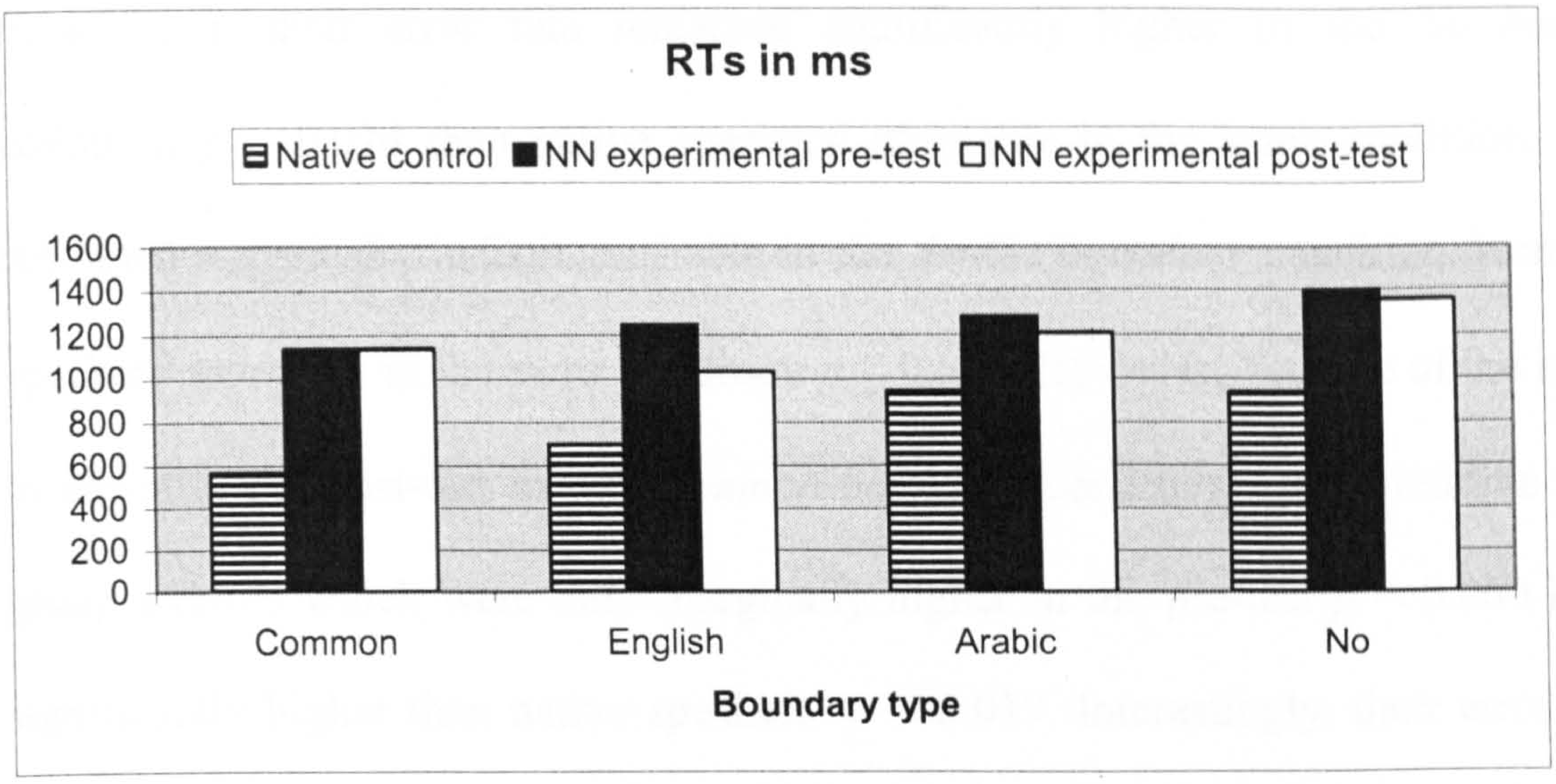

Figure 5.41 Native control group and non-native experimental group's mean RTs in ms in the pre-test and post-test in the Word Spotting Task in four different boundary conditions.

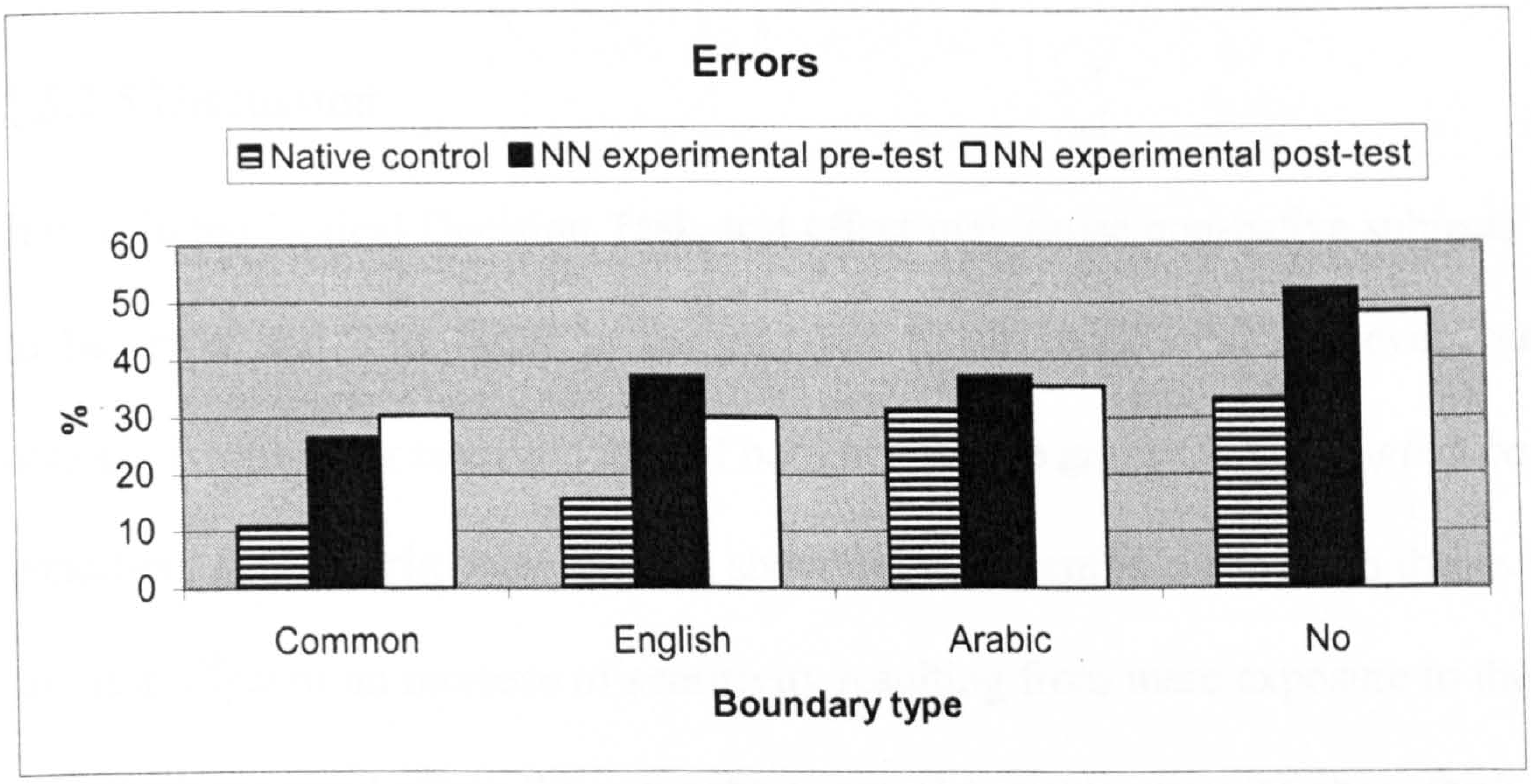

Figure 5.42 Native control group and non-native experimental group's mean percentage of errors in the pre-test and post-test in the Word Spotting Task in four different boundary conditions.

Two one-way ANOVAs were conducted to compare the native group's RTs (in the first ANOVA) and error rate (in the second) to those of the non-native experimental group's post-test's in all four conditions. Post hoc test results showed that the non-native experimental group's RTs remained significantly slower than native speakers' in all conditions (*Common boundary* condition p = 0.000; *English boundary* condition, p = 0.004; *Arabic boundary* condition p = 0.025; and *No boundary* condition p = 0.000).

In addition, their error rate remained significantly higher in the *No boundary* condition p = 0.024 than native speakers' error rate in the same condition. It also remained statistically indistinguishable in the *Arabic boundary* condition from native speakers' error rate in the same condition p = 0.620. However, because of the increase in errors in the post-test in the *Common boundary* condition the non-native control group's errors which were only marginally higher in the pre-test p = 0.060 became significantly higher than native speakers' p = 0.019. Interestingly, their error rate in the *English boundary* condition which was significantly higher in the pre-test p = 0.010 became only marginally higher in the post-test p = 0.084.

5.3.2.5 Discussion

Just as in the Lexical Decision Task, test effect may cause non-native subjects' errors to be fewer and RTs faster in the post-test in all conditions. However, our main concern is with error rates and RTs of both non-native groups in the *English boundary* condition. Did experimental subjects show improvement in a condition that is beyond any test effect or an increase of sensitivity resulting from mere exposure to these cues in naturalistic input where the non-native control subjects may also show this?

Non-native control. Post-test results showed that the non-native control showed no significant improvement neither in error rate or RTs in the post-test in all conditions. In addition, their RTs remained significantly slower than native speakers' in all conditions. Their error rates remained higher in the *English boundary* condition and the *No boundary* condition and statistically indistinguishable in the *Arabic boundary* condition. These results show that the naturalistic input which the control subjects were exposed to during the in-class activities and the weekly tasks, which were unrelated to phonotactics, did not improve their lexical segmentation ability using English phonotactic constraints.

The problem is that our control subjects, as is the case with all L2 learners, have L1 phonotactic constraints in place. The limited input non-immersion learners receive adds to the problem which seems to slow the acquisition of L2 phonotactic constraints. Awareness raising about these constraints and their use in lexical segmentation might be helpful. Given the little input and L1 influence, teaching phonotactics provides a shortcut to using L2 English constraints in lexical segmentation by directing learners' attention to the presence of these cues.

Non-native experimental. Our non-native experimental group's post-test results are promising and allow us to reject the part of null Hypothesis 14, which states that the experimental non-native subjects will not show any improvement towards their use of English phonotactic legality in lexical segmentation at post-test. The performance of our non-native experimental subjects improved. They became faster in detecting words in the *English Boundary* condition after the treatment. However, their RTs in the *English Boundary* condition were still significantly longer than native speakers'. This suggests that the experimental group's lexical segmentation ability was still not as automatic as the native speakers'. However, it is equally plausible that this difference is not because of a slower lexical segmentation, that is finding the boundary, but rather because of slower word recognition ability as confirmed by the Lexical Decision task. On the other hand, although their error rate did not become significantly lower in the post-test, it became almost statistically indistinguishable from native speakers'. Recall that, whereas some of the clusters in the list presented to the experimental group during the treatment were also illegal in Arabic, the only clusters from the list used to form a boundary constraint in the Word Spotting Task were those illegal solely in English (/dl/, /mr/, /bw/). That is why an improvement was expected to be in this condition only. These results can not be attributed to test or practice effects. First, training during the treatment period did not resemble the test. During training, subjects were asked to spot illegal clusters across word boundaries whereas the Word Spotting task required them to spot target words. Second, given the long time between the two tests, these results can not be a test effect. If they were, we would expect similar gains in the non-native control group's performance.[60] Similarly, if they were the result of a test effect, we would expect gains in the scores of all conditions. Clearly, this is not the case, as the experimental group's scores deteriorated in the post-test in the *Common Boundary* condition.

Another interesting result concerns our non-native subjects' performance in the *Arabic boundary* condition. In the post-test, both non-native groups still used Arabic phonotactic constraints to the same degree of speed and accuracy as they did in the pre-test. In other words, despite explicit teaching, the experimental subjects still transferred and used their L1 constraints which are not informative in English, suggesting that L1 cues are difficult to subdue when processing an L2.

5.4 Summary

In this chapter, results of the three subject groups in the three tasks were presented. The best way to summarise our results is to take each set in terms of answers to our

---

[60] However, although the two groups were comparable in terms of the length of the tasks they were set (see section 4.4.5) the experimental group may have the advantage of having more attention to input as they had to do the treatment requirement (i.e. underlining illegal clusters).

research questions. Research questions and their answers obtained from the results are presented below.

1- Are Arabic speaking EFL learners sensitive to the phonotactic constraints of English?

Evidence from the Non-word Rating Task suggests that they are. In this task non-native subjects in both groups were like native speakers in judging stimuli items in the *illegal in 1* condition (/dlɔɪθ/) to be less English-like than those in the *low probability* condition /ʃaiʒ/. In addition, the non-native control group made fewer errors when rating stimuli items in the *illegal in 1* condition than those in the *low probability* condition. These results allowed us to partially reject the null Hypothesis 3, which states that "non-native speakers will show no difference between their ratings or error rates in the *illegal in 2* condition, *illegal in 1* condition and the *low probability* condition"

2- Are Arabic speaking EFL learners sensitive to phonotactic probability in English?

Evidence from the Non-Word Rating Task and the Lexical Decision Task suggests that they are. In the Non-word rating Task, subjects in both non-native groups were like native speakers in judging stimuli items in the *high probability* condition (e.g. /saɪv/) more English-like than those in the *low probability* condition (e.g. /ʃaiʒ/). In addition, the non-native experimental group made more errors when judging stimuli items in the *high probability* condition than in the *low probability* condition (i.e. they wrongly classified some *high probability* non-words as real words). In the Lexical Decision task, our non-native subjects in both groups were like native speakers in deciding that non-words with low phonotactic probability (e.g. /ʃaiʒ/) are non-words more quickly and accurately than they did with those with high phonotactic probability. These results allowed us to reject both null Hypothesis 4, which states that "non-native speakers will show no difference between their ratings or error rates in the *high probability* condition and the *low probability* condition" and null Hypothesis 9, which states that "Non-native speakers will show no difference in RTs or error rates in rejecting stimuli items in the *high probability* condition and *low probability* condition"

3- Do Arabic speaking EFL learners use the phonotactic constraints of Arabic and English in lexical segmentation of running speech in English?

Evidence from the Word Spotting Task in the pre-test suggests that they use the phonotactic constraints of Arabic. Our non-native subjects in both groups were faster and more accurate in detecting a word aligned with an Arabic boundary constraint (e.g. ʒi:blɔ:rd) than when it was not. In addition, Evidence from the Word Spotting Task in the pre-test suggests that they use the phonotactic constraints of English but to a lesser degree than those of Arabic. Our non-native subjects in both groups were more accurate in detecting a word aligned with an English boundary constraint (e.g. ʒi:dlɔ:rd) than when it was not. This is also confirmed by the fact that subjects were on average fastest and most accurate when Arabic and English phonotactic constraints joined forces in the *Common Boundary* condition. These results allowed us to reject the null Hypothesis 13, which states that "Non-native speakers will show no difference in RTs or error rates in spotting words in the *No Boundary* condition on the one hand and the other three conditions on the other".

4- Can explicit teaching of English phonotactic constraints improve Arabic speaking EFL learners' ability in the lexical segmentation of English?

Evidence from the Word Spotting Task in the post-test suggests that it can. In the post-test, only the non-native experimental group became faster in detecting words in the *English Boundary* condition after the treatment. In addition, although their error rate did not become significantly lower in the post-test, it became statistically indistinguishable from native speakers'. These results allowed us to reject the part of the null Hypothesis 14, which states that "the experimental non-native subjects will not show any improvement towards their use of English phonotactic legality in lexical segmentation at post-test".

The next chapter discusses limitations of the present thesis. It also highlights pedagogical implications of the present findings and provides suggestions for future research.

Chapter 6: Conclusion and Pedagogical implications

6.1 Major findings

The present thesis has shown three major findings. First, it showed that like English native speakers, low-intermediate level EFL learners can show sensitivity to the phonotactic probability of English. In Chapter 5 I stated that EFL learners' sensitivity to the phonotactic probability of English suggests that this sensitivity is the result of abstract phonological information in memory that is independent of lexical representations of sound patterns and neighbourhood density effect. That is because the non-native lexicon and, more specifically neighbourhood density effects, are not expected to play the same role as in native speakers because of the different distribution of words in the native and non-native lexicons. Therefore, the findings of the current study are supportive of the proposal that phonotactic probability effects are unique and are not only subsumed by neighbourhood density effects (e.g. Vitevitch and Luce 1998; 1999; 2005).

The second major finding is that Arabic speakers low-intermediate EFL learners are sensitive to the phonotactic constraints of English and that they use phonotactic constraints of English and Arabic in the lexical segmentation of English connected speech. These results lend support to the Explicit Segmentation Models discussed in Chapter 3. They show that not only infants and adult native speakers but also FL learners use bottom-up cues to aid segmentation. This, however, does not mean that the listener does not use top-down cues when available. Lexical segmentation is a complex skill. For it to be successful the listener has to use all cues available, whether top-down lexical cues or bottom-up ones.

The third major finding is that explicit teaching of English phonotactic constraints can help improve Arabic speakers EFL learners' lexical segmentation ability using these constraints as segmentation cues in English running speech. Results of our experimental subjects in the Word Spotting Task in the post-test showed that they benefited from explicit teaching. Unlike the non-native control group, the experimental group became faster in detecting words in the *English Boundary* condition after the treatment. However, they were still slower than native speakers. Recall that in Chapter 2 I discussed that one of the ways explicit teaching can help is by improving the automaticity of learners performance. This finding suggests that controlled (non-automatic) / automatic processing should be viewed as a continuum rather than categorical. Accordingly, it seems that our subjects have improved a long that continuum. As discussed in Chapter 5, The fact that our experimental EFL learners still used their L1 constraints after the treatment in segmenting English connected speech suggests that L1 cues are difficult to subdue when processing an L2.

6.2 Limitations

Some limitations in the present study concern the implications of the findings rather than the methodology adopted. First, the advantage of using phonotactic constraints in connected speech may not always exist as not all word boundaries are marked by phonotactic cues, and some illegal clusters may be assimilated (e.g. bad guy /bægaɪ/). In addition, it is possible that assimilation can perhaps limit the value of phonotactic constraints as segmentation cues in connected speech. Using a computer speech recognizer it was shown by Harrington et al. (1989) that the faster the speech the more the variability in the signal limits the value of consonant sequences as segmentation cues.

Moreover, explicit teaching of phonotactic cues should be dealt with caution. Recall that during stage three (in-class practice) of the treatment period the activity used (spotting illegal clusters on-line) proved very difficult. In the beginning of this stage (week 5), students could spot a cluster only after at least the five next words were heard. However, students' performance seemed to have improved during the next weeks. At week 8, some students were able to spot a cluster after only the next word was presented, suggesting that such exercises might have overloaded students' processing capacity in the beginning.

In addition, the practicality of introducing phonotactic constraints in listening teaching materials may actually be reduced in ESL contexts. Unlike EFL classroom where students usually share the same language background, ESL classrooms are mixed environments involving students of different language backgrounds. Although it is possible for an ESL teacher to teach English phonotactic constraints to ESL students regardless of their L1 constraints, it is actually advisable for effective teaching of phonotactic constraints that the teacher knows the phonotactic constraints of the students' L1 to be able to concentrate on the L2 only constraints. While a simple contrastive analysis in an EFL classroom is possible, doing the same in an ESL classroom might be a daunting task.

6.3 Pedagogical implications

6.3.1 Extra-classroom exposure

What pedagogical implications could be drawn from these findings? The first obvious implication is the paramount importance of input. The pre-test results showed that all our low-intermediate EFL learners have developed a good level of sensitivity to both probability and constraints of English phonotactics. They also showed some exploitation of English phonotactic constraints when segmenting. Given the little chances of our EFL subjects to communicate with native speakers in that particular context and our subjects' statement that they have never been taught English phonotactic constraints before, it seems that they acquired this sensitivity based on the little naturalistic input they receive whether inside the classroom when set listening tasks or outside the classroom through the media. The first pedagogical implication therefore is that EFL students should be provided with as much naturalistic input as the time of the class would allow.

6.3.2 Awareness raising activities

The suggestion that input is important is anything but new. However, unlike previous similar suggestions of comprehension approaches (e.g. Krashen 1985; Ridgeway 2000), post-test findings point out to the importance of instruction in helping EFL learners make the most of the little input they receive. As was shown, only the experimental group have improved in their use of English phonotactic constraints as segmentation cues; they became faster and more accurate in using them. Therefore, the second pedagogical implication implied by the present study is that consciousness or awareness raising of the existence of these segmentation cues can result in a more automatic use of these cues in segmentation as shown by experimental subjects faster RTs in the Word Spotting task at post-test (Sharwood-Smith 1981).[61] Because

---

[61] I avoided stating that such training resulted in acquiring these cues. That is because to claim that this is the case a delayed test is needed (see limitations in Section 6.4 below).

automatic processes are less attention demanding, as discussed in Chapter 2, enhancing EFL learners' automaticity in lexical segmentation is supposed to help them pay more attention to higher-level (e.g. semantic, pragmatic and sociolinguistic) processes in listening comprehension (Segalowitz 2003). This way the horse is put before the cart (Norris 1995).

Awareness-raising exercises like those used in the present study are particularly important in EFL contexts given little naturalistic input EFL learners receive compared to ESL learners. Therefore, EFL learners need something to help them capitalise on the little input they receive.

EFL programs should therefore strike a balance between top-down and bottom-up training and aim to instruct students in using these bottom-up cues. Given little input, leaving EFL learners to their own devices would result in a delay in acquiring these cues. Awareness-raising instruction, as the current study has shown, was effective in drawing learners' attention to the effectiveness of inserting a boundary between phonemes in an illegal cluster.

However, as discussed in the previous section, phonotactic training exercise seems to have overloaded students' processing capacity in the beginning although they were low-intermediate level (i.e. not beginners). Therefore, such an exercise may not be suitable for beginning learners whose FL processing capacity is limited and who have not developed an appropriate level of FL knowledge.

6.3.3 Vocabulary teaching

The last pedagogical implication is inferred from the results which showed that our low intermediate EFL learners have developed a good level of sensitivity to the phonotactic probability of English. The fact that phonotactic probability can be acquired fairly quickly is also supported by studies, discussed in Chapter 3, which investigated the acquisition of artificial phonotactics (e.g. Chambers et al. 2003; Onishi et al 2002) and real language phonotactics (Gullberg et al. 2007). As was discussed in Chapter 3, being sensitive to the phonotactic probability of EFL may have some pedagogical implications for vocabulary learning. Recall that it was shown that native speakers show their sensitivity to the phonotactic probability of their L1 in different ways. Among these is the ease and speed of recognition, naming and recall of high phonotactic non-words (which resemble novel real words for EFL learners) (e.g. Thorn and Frankish 2005; Vitevitch et al. 1997; Vitevitch and Luce 1998; 1999; 2005). Consequently, presentation of new vocabulary items in vocabulary teaching materials may need to follow a specific graded order compatible with such findings.[62] In other words, high probability novel words could be presented earlier in vocabulary teaching programs as learners at beginning level might find them easier to recall and pronounce. After learners start showing some comfort in dealing with the EFL vocabulary items the probability of the new vocabulary items presented could be gradually changed till learners can be presented with vocabulary items with low phonotactic probability.

However, recall that recent evidence (Storkel et al. 2006) suggests that high phonotactic probability may have a disadvantageous effect when it comes to the speed

---

[62] A drawback here is communication needs which require introducing some common words regardless of their phonotactic probability.

of learning new vocabulary. As was discussed in Chapter 3, that is because high probability new items are deceptively similar to known items and therefore they may not be recognised as novel from a single exposure so learning may not be triggered. As a result a new representation may not be created (ibid). Given that our EFL subjects showed sensitivity to EFL phontactic probability and assuming that this argument can be extended to EFL vocabulary learning, then vocabulary teaching materials may aim to provide EFL learners, especially those having problems in learning (i.e. recognising that a new encountered word is novel and accordingly creating a new lexical representation) new EFL vocabulary items, with contextualised low probability new words which will supposedly be easier to learn. In sum, EFL learners who have difficulty in the pronunciation and recall of vocabulary items might be assisted by being provided with high probability words. On the other hand, those who have problems learning new vocabulary might be assisted by being provided with contextualised low probability words more often than high probability ones.

6.4 Suggestions for further research

Areas for further research can actually be discussed here. First, a delayed post-test could provide some information about the source of improvement. Is it a temporary training effect? Or is it a lasting effect such that learners will notice illegal clusters in subsequent input thereby helping them acquire these cues and accelerating automisation of their lexical segmentation ability? The researcher's first research project after finishing this thesis is to investigate this question by testing the same subjects.

The effect of longer training should also be investigated. The treatment in the current study lasted about eight weeks with an average of one hour and a half of instruction a week. Although experimental subjects became significantly faster in the *English Boundary* condition in the post-test, their RTs were still significantly longer than native speakers'. Longer training would show us if subjects' performance could approach that of native speakers. Also, examining the effectiveness of phonotactic cues in segmentation by using tasks which include longer stretches of naturalistic connected speech would show more clearly if phonotactic cues are really helpful in segmenting connected speech. One good but rather difficult to design way to improve upon the current methodology is to have an integrative test (e.g. dictation or a comprehension test) where phonotactic cues are integrated. An example of a sentence in this test would look like the sentence in (1) where underlined between words segments form illegal clusters.

(1) pick boring boys and leave them out.

This is not counter to the argument (see Chapter 4) that integrative tests alone are not good for measuring the effect of teaching bottom-up cues; rather it is suggested here that a modified integrative test (like the one proposed above) should be used alongside the Word Spotting Task which can directly measure the improvement in the targeted skill.

Future research should also target other cues. Field (2001; 2003) for instance has suggested that ESL learners could be trained to use stress as a segmentation cue. It seems that just like English native speakers (Cutler and Norris, 1988) it will be safe for ESL learners to assume word boundaries upon hearing stressed syllables. As discussed in Chapter 3, the beginning of large percentage of English words could be identified this way (Cutler and Carter, 1987). It will be interesting to find out that training ESL learners whose L1 use a different prosody-based segmentation strategy to use stressed syllables as a segmentation cue in English is possible. That is because, unlike phonotactic constraints, evidence suggests that implicitly learning an L2-specific prosodic segmentation strategy is difficult (Cutler et al. 1992).

Another area which merits further research is disassociating the effect of the two processes involved in the Word Spotting Task, namely segmentation and word recognition. Recall that in the post-test the experimental group's RTs in the *English Boundary* condition were still significantly longer than native speakers'. One plausible explanation is that the experimental group's lexical segmentation ability was still not as automatic as the native speakers'. However, it is equally plausible that this difference is not because of a slower lexical segmentation ability, i.e. finding the boundary, but rather because of a slower word recognition ability, as confirmed by the Lexical Decision task, or both.[63] Future research might therefore try to disassociate this effect by presenting the same target words in the word spotting task in a lexical decision one. This way we can find out which of the two processes, segmentation or word recognition is responsible for the delay in responses in the Word Spotting task.

Future research might also examine the viability of integrating Phonotactic training with ear training and pronunciation teaching programs. Recall that our non-native speakers transferred their L1 phonotactic constraints when listening to English even when these phonotactics were not helpful in EFL. We need to know if, with appropriate practice, EFL learners could be trained to perceive and pronounce without epenthetic vowels the legal English consonant clusters that are illegal in their native

---

[63] In the lexical decision task our non-native subjects were significantly slower than native speakers in rejecting non-words suggesting lack of automaticity in their L2 word recognition ability.

language. If successful, such training can potentially play a role in limiting L1 transfer in segmentation.

6.5 Conclusion

Several calls have been made to enhance L2 learners' bottom-up listening skills (e.g. Cauldwell 1996; Field 2003; Hulstijn 2001). The language specificity of the cues used in bottom-up listening skills such as lexical segmentation is a source of listening difficulty for L2 learners (Weber and Cutler, 2006). Using a well controlled design the current study showed that with appropriate training EFL learners could be taught to use the EFL specific phonotactic cues faster and more accurately in lexical segmentation.

A final remark should be stressed. Predominantly using a top-down approach, we are only teaching learners how to listen. However, if we teach them how to automate the bottom-up skills we are not only teaching them how to listen but also how to learn the L2 through listening. Listening materials' writers and listening teachers should bear that in mind and act accordingly. One way to do that, as this study has shown, is by providing students with tasks that can help them automate using L2-specific cues for those bottom-up skills they have already developed in their L1.

References

Abercrombie, D. (1964). A phonetician's view of verse structure. *Linguistics* 6: 5-13.

Akita, M. (2005). The effectiveness of a prosody-oriented approach in L2 perception and production. In: *BUCLD 29 Proceedings*: A. Brugos, M. Clark-Cotton and S. Ha (eds.). Cascadilla Press, Somerville, MA.

Altenberg, E. (2005a) The perception of word boundaries in a second language. *Second Language Research* 21 (4): 325-358.

Altenberg, E. (2005b) The judgement, perception and production of consonant clusters in a second language. *IRAL* 43: 53-80.

Altenberg, E. and Cairns, H. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behaviour* 22: 174–188.

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review* 89 (40): 369-406.

Anderson, J. (1983) *The Architecture of Cognition.* Harvard University Press, Cambridge, MA.

Anderson, J. (1995). *Cognitive Psychology and its Implications*. 4th edition. W. H. Freeman and Company, New York.

Anderson, J. and Lebiere, C. (1998). *The Atomic Components of Thought.* Lawrence Erlbaum Associates, Mahwah, NJ.

Anderson, J. and Fincham, J. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20: 1322–1340.

Azar, B. (1996) *Basic English Grammar.* 2nd edition. Prentice Hall Regents, Upper Saddle River, NJ.

Baayen, H., Piepenbrock, R. and van Rijn, H. (1993). *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. (1986). *Working Memory.* Oxford University Press, Oxford.

Baddeley, A. (1990). *Working Memory: Theory and Practice.* Lawrence Erlbaum Associates, London.

Baddeley, A. (1999) *Essentials of Human Memory.* Psychology Press Ltd, Sussex.

Baddeley, A. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences* 4 (11): 417-423.

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

Baddeley, A. and Hitch, G. (1974). Working memory. In: *Recent Advances in Learning and Motivation: Vol. 8,* G. Bower (ed.). Academic Press, New York.

Bailey, T. and Hahn, U. (2001). Determinants of wordlikeness: phonotactics or lexical neighbourhoods? *Journal of Memory and Language* 44: 568-591.

Bard, E., Shillcock, R. and Altmann, G. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics* 44: 395-408.

Baumann, H., Nagengast, J. and Klaas, G. (1993). New experimental setup (NESU). Max Planck Institute for Psycholinguistics, Ms. Nijmegen.

Bernstein, N. (1982). *Acoustic Study of Mothers' Speech to Language-learning Children: An Analysis of Vowel Articulatory Characteristics.* PhD Thesis. Boston University, USA.

Bialystok, E. (1994). Representations and ways of knowing: three issues in second language acquisition. In: *Implicit and Explicit Learning of Languages*, N. Ellis (ed.). Academic Press, London. pp. 549-569.

Bond, Z. (1971). Units in speech perception. *Ohio State University Working Papers in Linguistics* 9: 1-112.

Bowen, J. (1976). Current research on an integrative test of English grammar. *RELC Journal* 7: 30-37.

Brent, M. and Cartwright, T. (1996). Distributional regularity and phonotactics are useful for segmentation. *Cognition* 61: 93-125.

Broselow, E. and Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35–59

Brown, D. (1994) *Teaching by Principles: An Interactive Approach to Language Pedagogy.* Prentice-Hall International (UK) Limited, London.

Brown, G. (1990). *Listening to Spoken English.* Longman, London.

Brown, J. and Hilfetry, A. (1987). The effectiveness of teaching reduced forms for listening comprehension. *RELC Journal* 17 (2): 59-70.

Brown, R. and Hildum, D. (1956). Expectancy and the perception of syllables. *Language* 32: 411-419.

Buck, G. (2001). Assessing Listening. Cambridge University Press, Cambridge.

Cairns, P., Shillcock, R., Chater, N. and Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology* 33: 111-153.

Call, M. (1985). Auditory short-term memory, listening comprehension, and the Input Hypothesis. *TESOL Quarterly* 19: 765-781.

Carroll, D. (1986). *Psychology of Language*. Wadsworth, Inc., Belmont, CA.

Carroll, S. (2001) *Input and Evidence: The Raw Material of Second Language Acquisition*. John Benjamins publishing company, Philadelphia.

Carroll, S. (2004) Segmentation: Learning how to 'hear words' in the L2 Speech Stream. *Transactions of the Philological Society* 102 (2): 227–254.

Cauldwell, R. (1996). Direct encounters with fast speech on CD- audio to teach listening. *System* 24 (4): 521-528.

Chambers, K., Onishi, K. and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87: B69–B77.

Chamot, A. (1995). Learning strategies and listening comprehension. In: *A Guide for the Teaching of Second Language Listening*, D. Mendelsohn and J. Rubin (eds.) Dominie Press, San Diego. pp. 13–30.

Chamot, A. and O'Malley, J. (1994). Language learner and learning strategies. In: *Implicit and Explicit Learning of Languages*, N. Ellis (ed.). Academic Press, London. pp. 371-92.

Chomsky, N. (1975) *Reflections on Language*. Pantheon, New York

Christiansen, M., Allen, J. and Seidenberg, M. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes* 13: 221-268.

Christie, W. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America* 55 (4): 819-821.

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition* 25: 35-69.

Clark, H. and Clark, E. (1977). Psychology and language. Harcourt Brace Jovanovitch, New York.

Cole, R. and Jakimik, J. (1980). A model of speech perception. In: *Perception and Production of Fluent Speech*, R. Cole (ed.), Lawrence Erlbaum Associates, Hillsdale, NJ. pp. 133-163.

Coleman, J. and Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. *Computational Phonology* 3: 49-56.

Cooney, T., Cleary, C. and Holden, B. (2003). *Top-up Listening 3*. ABAX Ltd, San Francisco.

Corder, S. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching* 5: 161-170.

Craik, F. and Lochhart, R. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour* 1: 671-684.

Crain, S. (1993). Language acquisition in the absence of experience. In: *Language Acquisition: Core Readings*, P. Bloom (ed.) Harvester Wheatsheaf, London.

Crystal, D. (1980). *A First Dictionary of Linguistics and Phonetics*. Westview Press, Boulder, Colo.

Crystal, T. and House, A. (1988). Segmental durations in connected speech signals: Current results. *Journal of Acquistical Society of America* 83: 1553-1573.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In: *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, G. Altmann (ed.). MIT Press, Cambridge, MA. pp. 105-121.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua* 92: 81-104

Cutler, A. (1996). Prosody and the word boundary problem: In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, J. Morgan and K. Demuth (eds.). Erlbaum, Mahwah, NJ. pp. 87–99.

Cutler, A. and Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture mis-perception. *Journal of Memory and Language* 31: 218–236.

Cutler, A. and Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2: 133-142.

Cutler, A. and Foss, D. (1977). On the role of sentence stress in sentence processing. *Language and Speech* 20: 1-10.

Cutler, J., Mehler, J., Norris, D. and Seguí, J. (1986) The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385-400.

Cutler, A., Mehler, J., Norris, D. and Seguí, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology* 19: 141–77.

Cutler, J., Mehler, J., Norris, D. and Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* 24: 381–410.

Cutler, A. and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception* *and Performance* 14: 113–121.

Cutler, A. and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33: 824-844.

Dalton, C. and Seidlhofer, B. (1994) *Pronunciation.* Oxford University Press, Oxford.

Davenport, M. and Hannahs, S. J. (1998). Introducing Phonetics and Phonology. Arnold, London.

Davis, M. (2000). *Lexical Segmentation in Spoken Word Recognition.* PhD Thesis, University of London, UK.

DeKeyser, R. (2003). Implicit and explicit learning. In: *The Handbook of Second Language Acquisition*, C. Doughty and M. Long (eds.). Blackwell Publishing Ltd, Oxford. pp. 313-348.

Dell, G., Reed, K., Adams, D. and Meyer, A. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1355–1367.

Doughty, C. (2003). Instructed SLA: Constraints, Compensation, and Enhancement. In: *The Handbook of Second Language Acquisition*, C. Doughty and M. Long (eds.). Blackwell Publishing Ltd, Oxford. pp. 256-310.

Dumay, N., Content, A. and Frauenfelder, U. (1999). Acoustic–phonetic cues to word boundary location: Evidence from word-spotting. In: *Proceedings of the 14th International Congress of Phonetic Sciences.* University of California, Berkeley, CA. pp. 281-284.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25 (6): 1568–1578.

Dupoux, E., Pallier, C., Kakehi, K. and Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16: 491–505.

Eckman, F. (1987). Markedness and the contrastive analysis hypothesis. In: *Interlanguage phonology: The acquisition of a second language sound system*, G. Ioup and S. Weinberger (eds.). Newbury House Publishers, Cambridge.

Eimas, P., Hornstein, S. and Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language* 29 (2): 160-180.

Ellis, N. (1994a). Implicit and explicit language learning: an overview. In: *Implicit and Explicit Learning of Languages*, (ed.). Academic Press, London. pp. 1-31.

Ellis, R. (1994b). A theory of instructed second language acquisition. In: *Implicit and Explicit Learning of Languages*, N. Ellis (ed.). Academic Press, London. pp. 79-114.

Estes, W. (1973) Phonemic coding and rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behaviour* 12: 360-372.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. Infant Behaviour and Development 8: 181-195.

Field, J. (1997). Monitoring for breakdown of understanding. In: *Teachers Develop Teachers Research, Vol. 2*, J. Field, A. Graham, E. Griffiths and K. Head (eds.). Whitstable, IATEFL.

Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal* 52: 110–118.

Field, J. (1999). "Bottom-up" and "Top-down". *ELT Journal* 53 (4): 338-339.

Field, J. (2001) *Lexical Segmentation in First and Foreign Language Acquisition.* PhD thesis. University of Cambridge, UK.

Field, J. (2003) Promoting perception: lexical segmentation in L2 listening. *ELT Journal* 57 (4): 325-234.

Field, J. (2004) An insight into listeners' problems: too much bottom-up or too much top-down? *System* 32 (3): 363-377.

Filología Inglesa, D. (non-dated). EFL university students' cognitive processing of spoken academic discourse as evidenced by lecture notes, http://www.publicacions.ub.es/revistes/bells12/PDF/art14.pdf, 21 May 2006.

Feyten, C. (1991). The Power of Listening Ability: An Overlooked Dimension in Language Acquisition. *The Modern Language Journal* 75: 173-80.

Flowerdew, J., and Miller, L. (2005). *Second language listening: Theory and practice.* Cambridge University Press, Cambridge.

Fodor, J. (1983). *The Modularity of Mind.* : MIT Press, Cambridge, MA.

Forster, K. and Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12: 627- 635.

Foss, D. and Blank, M. (1980). Identifying the speech codes. *Cognitive Psychology* 12:1-31.

Frauenfelder, U. and Peeters, G. (1990). On lexical segmentation in TRACE: An exercise in simulation. In: *Cognitive Models of Speech Processing: Psycholinguistic and computational perspectives,* G. Altmann (ed.). MIT Press, Cambridge, MA.

Frisch, S., Large, N. and Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on processing non-words. *Journal of Memory and Language* 42: 481-96.

Ganong, W. (1980). Phonetic categorisation in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6: 110-125.

Gass, S. (1990). Second and foreign language learning: Same, different or none of the above? In: *Second language acquisition,* B. VanPatten and J. Lee (eds.). Multilingual Matters, Clevedon, UK. pp. 34-44.

Gass, S. (1991). Grammar instruction, selective attention, and learning. In: *Foreign/Second Language Pedagogy Research,* R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith and M. Swain (eds.). Multilingual Matters, Clevedon, UK. pp. 124-141.

Gass, S. (2003) Input and interaction. In: *The Handbook of Second Language Acquisition,* C. Doughty and M. Long (eds.). Blackwell Publishing Ltd, Oxford. pp. 224-255.

Gass, S. and Madden, C. (1985). Introduction. In: *Input in Second Language Acquisition* (eds.). Newbury House Publishers, Cambridge.

Gathercole, S. and Baddeley, A. (1989) Evaluation of the role of STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language* 28: 200-213.

Gathercole, S. and Baddeley, A. (1993). *Working Memory and Language.* Lawrence Erlbaum Associates, Hillsdale, New York.

Gaygen, D. (1997). *The Effects of Probabilistic Phonotactics on the Segmentation of Continuous Speech.* PhD Dissertation, State University of New York at Buffalo, USA.

Gaygen, D. and Luce, P. (2002). Troughs and busts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech. In: *University at Buffalo Working Papers on Language and Perception, Vol. 1,* C. McLennan, P. Luce, G. Mauner, and J. Charles-Luce (eds.). pp. 496 – 548, http://66.102.9.104/search?q=cache:ZyGl4EK5E7kJ:languageperception.buffalo.edu/lswp/1.9%2520G%26L.pdf+Gaygen,+D.+and+Luce,+P.+(2002).&hl=en&ct=clnk&cd=9&gl=uk, 5th April 2005.

Geigerich, H. (1992). *English Phonology: An Introduction.* Cambridge University Press, Cambridge.

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing* 19: 133–167.

Gimson, A. (2001) *Gimson's Pronunciation of English.* Arnold, London.

Glen, E. (1989) A content analysis of fifty definitions of listening. *Journal of the International Listening Association* 3: 21-31.

Goh, C. (1997). Metacognitive awareness and second language listeners. ELT Journal 51(4): 361-369.

Goh, C. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research* 2: 124–147.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System* 28: 55-75.

Goh, C. (2002) Exploring listening comprehension tactics and their interaction patterns. *System* 30: 185–206.

Goldinger, S. (1996). Auditory lexical decision. *Language and* Cognitive Processes 11: 559–567.

Goodman, J. and Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language* 27: 684–698.

Graham, Z. (2006). Listening comprehension: The learners' perspective. *System* 34: 165-182.

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning* 40: 311-336.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly* 26: 385-391.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics* 38: 299–310.

Grosjean, F. (2001). The bilingual's language modes. In: *One Mind, Two Languages*, J. Nicol (ed.) Blackwell, Oxford. pp.1–22.

Gullberg, M., Dimroth, C. and Roberts, L. (2007) What can adults learn about word forms in an unknown language after minimal exposure? A paper presented at the 17th European Second Language Association Conference, Newcastle University, UK.

Hansen, C. and Jensen, C. (1994). Evaluating lecture comprehension. In: *Academic Listening,* J. Flowerdew (ed.). Cambridge University Press, Cambridge. pp. 241–268

Harrington, J., Watson, G., Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language* 3: 367-382.

Henrichsen, L. (1984). Sandhi-variation: A filter of input for learner of ESL. *Language Learning* 34 (3): 103-126.

Herron, C., Cole, S., York, H., and Linden, P. (1998). A comparison study of student retention of foreign language video: Declarative versus interrogative advance organizers. *The Modern Language Journal* 82: 237–247.

Herron, C ., Hanley, J. and Cole, S. (1995). A comparison study of two advance organizers for introducing beginning foreign language students to video. *The Modern Language Journal* 79 (3): 387-395.

Hildyard, A. and Olson, D. (1982). On the comprehension and memory of oral versus written discourse. In: *Spoken and Written language,* D Tannen (ed.). NJ Ablex, Norwood. pp 19-32.

Hohne, E. and Jusczyk, P. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics* 56: 613-623.

Hollich, G., Jusczyk, P. and Luce, P. (2001). Infants' memory for similar sounding words: phonetic false memories. Paper presented to the Society for Research on Child Development. Minneapolis, MN.

Hulstijn, J. (1990). A comparison between the information-processing and the analysis/control approaches to language learning. *Applied Linguistics* 11: 30–45.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In: *Cognition and Second Language Instruction,* P. Robinson (ed.). Cambridge University Press, Cambridge. pp. 258–286.

Ito, Y. (2001). Effect of reduced forms on ESL learners' input-intake process. *Second Language Studies* 20 (1): 99-124.

Johnson, K. (1996). *Language Teaching and Skill Learning.* Blackwell, Oxford.

Johnson, E. and Jusczyk, P. (2001). Word segmentation by 8-montholds: When speech cues count more than statistics. *Journal of Memory and Language,* 44 (4) 548–567.

Jones, L. and Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal* 86: 546–561.

Jusczyk, P. (1997) *The Discovery of Spoken Language.* MIT Press, Cambridge MA.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science* 3: 323-328.

Jusczyk, P., Cutler, A. and Redanz, N. (1993a). Preference for the predominant stress patterns of English words. *Child Development* 64: 675-687.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V. and Jusczyk, A. (1993b). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32, 402–420.

Jusczyk, P., Luce, P. and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33: 630–645.

Jusczyk, P., Hohne E. and Bauman, A. (1999a). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics* 61(8): 1465-1476.

Jusczyk, P., Houston, D. and Newsome, M. (1999b). The beginnings of word segmentation in English learning infants. *Cognitive Psychology* 39: 159-207.

Kellerman, S. (1990) Lip service: The contribution of the visual modality to speech perception and its relevance to the teaching and testing of foreign language listening comprehension. *Applied Linguistics* 11 (3): 272-290.

Kemler-Nelson, D., Jusczyk, P., Mandel, D., Myers, J., Turk, A. and Gerken, L. (1995). The headturn Preference Procedure for testing auditory perception. *Infant Behavior and Development* 18: 111–116.

Klatt, D. (1989). Review of selected models of speech perception. In: *Lexical Representation and Process*, W. Marslen-Wilson. MIT Press, Cambridge. pp. 169-226.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning.* Pergamon, Oxford.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition.* Pergamon, Oxford.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications.* Longman, London.

Krashen, S. (1994). The input hypothesis and its rivals. In: *Implicit and Explicit Learning of Languages*, N. Ellis (ed.). Academic Press, London. pp. 45-77.

Krashen, S. (1996). The case for Narrow Listening. *System* 24: 97–100.

Kucera, H. and Francis, W (1967). *Computational Analysis of Present-Day American English.* Brown University Press, Providence.

Kumaravadivelu, B. (1994). Intake factors and intake processes in adult language learning. *Applied Language Learning* 5: 33-71.

Lamel, L. and Zue, V. (1984) Properties of consonant sequences within and across word boundaries. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. pp. 296-299.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. Supplement to *Phonetica* 5 : 1–54.

Leigh, P. and J. Charles-Luce (2002) Developmental effects of phonotactic probability and lexicality on duration in speech production. In: *University at Buffalo Working Papers on Language and Perception, Vol. 1,* C. McLennan, P. Luce, G. Mauner, and J. Charles-Luce (eds.). pp. 603-653.

Levelt, W. (1978). Skill theory and language teaching. *Studies in Second Language Acquisition* 1: 53–70.

Levelt, W. (1989). *Speaking: From Intention to Articulation.* : MIT Press, Cambridge, MA.

Lively, S., Pisoni, D. and Goldinger, S. (1994). Spoken word recognition: research and theory. In: *Handbook of Psycholinguistics*, M. Gernsbacher (ed.). Academic Press, San Diego, CA. pp. 265-301.

Loftus, G. and Loftus, E. (1976). *Human Memory: The Processing of Information.* Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

Long, D. (1990). What you don't know can't help you complete. *Studies in Second Language Acquisition, 12,* 65–80.

Long, M. (1996). The role of linguistic environment in second language acquisition. In: *Handbook of Second Language Acquisition*, W. Ritchie and T. Bahtia (eds.). Academic Press, New York. pp. 413-468.

Luce, P. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics* 39: 155-158.

Luce, P. and Pisoni, D. (1998). Recognizing spoken words: The neighbourhood activation model. *Ear and Hearing* 19:1–36.

Luce, A. and Large, N. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes* 16 : 565–581

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics* 18: 3–19.

Lynch, T. (2002). Listening: Questions of level. In: *Oxford Handbook of Applied linguistics*, R. Kaplan (ed.). Oxford University Press, Oxford. pp. 39–48.

Marcus S. and Frauenfelder, U. (1985). Word recognition - uniqueness or deviation? A theoretical note. *Language and Cognitive Process* 1: 163-169.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature* 244: 522-523.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25: 71-102.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In: *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, G. Altmann (ed.). MIT Press, Cambridge, MA. pp. 148-172.

Marslen-Wilson, W. and Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology* 10: 29-63.

Massaro, D. and Cohen, M. (1983). Phonological context in speech perception. *Perception and Psychophysics* 34: 338-348.

Matthews, J. and Brown, C. (2004). When intake exceeds input: Language specific perceptual illusions induced by L1 prosodic constraints. *International Journal of Bilingualism* 8 (1): 5-27.

Mattys, S. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin and Review* 4 (3): 310–329.

Mattys, S., Jusczyk, P., Luce, P. and Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* 38: 465-494.

Mattys, S. and Jusczyk, P. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78: 91-121.

McClelland, J. and Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1-86.

McQueen, J. (1996). Word-spotting. *Language and Cognitive Processes* 11: 695–699.

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language* 39: 21-46.

McQueen, J., Cutler, A., Briscoe, T. and Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 10: 309-331.

McQueen, J. and Cutler, A. (1997). Cognitive processes in speech perception. In: *The Handbook of Phonetic Sciences*, W. Hardcastle and J. Laver (eds.). Blackwell, Oxford. pp. 566–585.

McQueen, J., Norris, D. and Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 621–638.

McLaughlin, B. (1987). *Theories of Second language Learning.* Edward Arnold, London.

McLaughlin, B., Rossman, T. and McLeod, B. (1983). second language learning: an information processing perspective. *Language Learning* 33 (1): 135-57.

McLaughlin, B. and Heredia, R. (1996). Information-processing to research on second language acquisition and use. In: *Handbook of Second Language Acquisition*, W. Ritchie and T. Bahtia (eds.). Academic Press, New York. pp. 213-28.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In: *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, G. Altmann (ed.). MIT Press, Cambridge, MA. pp. 236-262.

Mendelsohn, D. (1994). Learning to Listen : A Strategy-based Approach for the Second Language Learner. Dominie press, San Diego.

Mendelsohn, D. (1995). Applying learning strategies in the second/foreign language listening comprehension lesson. In: *A Guide for the Teaching of Second Language Listening*, D. Mendelsohn and J. Rubin (eds.) Dominie Press, San Diego.

Mendelsohn, D. (1998). Teaching listening. *Annual Review of Applied Linguistics* 18: 81–101.

Messer, S. (1967). Implicit phonology in children. *Journal of Verbal Learning and Verbal Behaviour* 6: 609-613.

Morgan, J. and Saffran, J. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development 66: 911-936.

Morley, J. (2001). Aural comprehension instruction: principles and practices. In: Teaching English as a Second or Foreign Language, 3rd Edition, M. Celce-Murcia (ed.). Heinle and Heinle, Boston. pp. 69-85.

Mueller, G. (1980). Visual Contextual Cues and Listening Comprehension: An Experiment. *The Modern Language Journal* 64 (3): 335-340.

Nakatani, L. and Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of Acquistical Society of America* 62 (3): 714-719.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52(3): 189-234.

Norris, D., McQueen, J. and Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23(3): 299-370.

Norris, D., McQueen, J., Cutler, A. and Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. Cognition and Psychology 34: 191–243.

Norris, R. (1995). Teaching reduced forms: putting the horse before the cart. *English Teaching Forum* 33: 47-50, http://www2.gol.com/users/norris/reduced3.html, 2nd March 2006.

Oakeshott-Taylor, A. (1977). Dictation as a test of listening comprehension. In: *Hőrverständnis im Fremdsprachenunterricht. Listening Comprehension in Foreign Language Teaching,* R. Dirven (ed.). Scriptor, Kronberg. PP. 131-154.

Oller, J. (1979). *Language Tests at School.* Longman, London.

O'Malley, J. (1987). The effects of training in the use of learning strategies on learning English as a second language. In: *Learner Strategies in Language Learning,* A. Wenden and J. Rubin (eds.). Prentice Hall International, Englewood Cliffs, NJ. pp. 133-144.

O'Malley, J., Chamot, A. and Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied linguistics* 10 (4): 418-437.

O'Malley, M. and Chamot, A. (1990). *Learning Strategies in Second Language Acquisition.* Cambridge University Press, Cambridge.

Onishi, K., Chambers, K. and Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition* 83:13-23.

Otake, T., Hatano, G., Cutler, A. and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32: 358–378.

Oxford Advanced Learners Dictionary (2000). Oxford University Press, Oxford.

Oxford, R. (1993). Research update on teaching L2 listening. *System,* 21(2): 205-211.

Papagno, C., Valentine, T. and Baddeley, A. (1991). phonological short-term memory and foreign language vocabulary learning. *Journal of Memory and Language* 30: 331-347.

Peterson, L. and Johnson, S. (1971). Some effects of minimising articulation on short-term retention. *Journal of Verbal Learning and Verbal Behaviour* 10: 346-354.

Peterson, P. (1991). A synthesis of methods for interactive listening. In: *Teaching English as a Second or Foreign Language,* 2nd Edition, M. Celce-Murcia (ed.) Newbury House, New York. pp. 11-22.

Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics* 60: 941–951.

Potter, R., Kopp, G. and Kopp, H. (1966). Visible Speech. Dover publications, http://www.geocities.com/agihard/mohl/mohl_language_substance.html#Visible%20Speech, 23rd September 2005.

Rankin, P. (1930) Listening Ability: Its Importance, Measurement, and Development. *Chicago Schools Journal* 12:177-179.

Richards, J. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly* 17 (2): 219-40.

Ridgway, T. (2000). Listening strategies_ I beg your pardon. *ELT Journal* 54 (2): 179-85.

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning* 45: 283–331.

Rosa, M. (2002). Don'cha know? A survey of ESL teachers' perspectives on reduced forms instruction. *Second Language Studies* 21 (1): 49-78.

Rost, M. (2001). Teaching and Researching Listening. Longman, London.

Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal* 78 (2): 199-221.

Saffran, J., Aslin, R. and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.

Samuel, A. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 7: 1124–1131.

Samuel, A. (1991). A further examination of attentional effects in the phonemic restoration illusion. *Quarterly Journal of Experimental Psychology* 43A: 679–699.

Samuel, A. (1996a). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 125: 28–51.

Samuel, A. (1996b). Phoneme restoration. *Language and Cognitive Processes* 11: 647–653.

Sanders, L., Neville, H. and Woldorff, M. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research* 45: 519-530.

Savin, H. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America* 35: 200-206.

Schmidt, R. (1983) Interaction, acculturation and the acquisition of communicative competence. In: *Sociolinguistics and Second Language Acquisition,* N. Wolfson and E. Judd (eds.). Newbury House, Rowley, Mass. pp. 137-74.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11: 128-158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14: 357–385.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In: *Implicit and Explicit Learning of Languages*, N. Ellis (ed.). Academic Press, London. pp. 165-210.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In: *Attention and Awareness in Foreign Language Learning,* R. Schmidt (ed.), University of Hawaii press, Honolulu. pp. 1-65.

Schmidt, R. (2001). Attention. In: *Cognition and Second Language Instruction*, P. Robinson (ed.) Cambridge University Press, Cambridge. pp. 3–32.

Schmidt, R. and S. Frota (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In: *Talking to learn: Conversation in Second Language Acquisition*, R. Day (ed.). Newbury House, Rowley, MA. pp. 237-326.

Schneider, W. and Fisk, A. (1983). Attentional theory and mechanisms for skilled performance. In: *Memory and Control of Action*, R. Magill (ed.). North-Holland Publishing Company, NY. pp.119-143.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behaviour. *Studies in Second Language Acquisition* 15 (2): 147-163.

Segalowitz, N. and Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics* 14: 369–385.

Segalowitz, N. (2003). Automaticity and second languages. In: *The Handbook of Second Language Acquisition*, C. Doughty and M. Long (eds.). Blackwell Publishing Ltd, Oxford. pp.382-408.

Service, E. (1992). Phonology, working memory, and foreign language learning. *The Quarterly Journal of Experimental Psychology*, 45A (1): 21-50

Sharwood Smith, M. (1981). Consciousness-raising and second language acquisition. *Applied Linguistics* 2 (2): 159-168.

Sharwood Smith, M. (1986). Comprehension versus acquisition: two ways of processing input. *Applied Linguistics* 7: 239-56.

Sheerin, J. (1987). Listening comprehension: Teaching or testing? *ELT Journal* 4: 126–131.

Shiffrin, R. and Schneider, W. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending and a general theory. *Psychological Review* 84: 127-190.

Shiffrin, R. and Schneider, W. (1984) Automatic and control processing revisited. *Psychological Review* 91(2): 269-76.

Shohamy, E. and Inbar, O. (1991) Validation of listening comprehension tests: The effect of text and question type. *Language Testing* 8(1): 23-40.

Siqueland, E. and DeLucia, C. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science* 165: 1144–1146.

Stanley, J. (1978). Teaching listening comprehension: An interim report on a project to use uncontrolled language data as a source material for training foreign students in listening comprehension. *TESOL Quarterly* 12 (3): 285-295.

Strokel, H., Armbrüster, J. and Hogan, T. (2006). Differentiating Phonotactic Probability and Neighborhood Density in Adult Word Learning. *Journal of Speech, Language, and Hearing Research* 49: 1175–1192.

Tarone, E. (1980). Communication, strategies, foreigner talk and repair in interlanguage. *Language Learning* 30: 417-431.

Tauroza, S. and Allison, D. (1990). Speech rates in British English. *Applied Linguistics* 11(1): 90 –105.

Thompson, I. and Rubin, J. (1996). Can strategy instruction improve listening comprehension? *Foreign Language Annals* 29 (3): 331-342.

Thorn, A. and Frankish, C. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. Journal of Experimental Psychology: Learning, Memory, and Cognition 31: 729–735.

Tsui, A. and Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics* 19 (4): 432-451.

Vainikka, A., and Young-Scholten, M. (2007). The role of literacy in the development of L2 morph-syntax from an organic grammar perspective. In: *Low-Educated Second Language and Literacy Acquisition, Research, Policy and Practice,* N. Faux (ed.). The Literacy Institute, Richmond, VA.

Vandergrift, L. (1996). The listening comprehension strategies of core French high school students. *Canadian Modern Language Review* 52: 200-223.

Vandergrift, L. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *The Modern Language Journal* 81(4): 494-505.

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal* 53(3): 168-176.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics* 24: 3-25.

van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech. *Perception and Psychophysics* 63: 811–823.

Van de Weijer, J. (1998). *Language Input for Word Discovery*. PhD Thesis, University of Nijmegen, Nijmegen.

Vitevitch, M. (2005) *Personal communication*, Kansas University, USA.

Vitevitch, M., Luce, P., Charles-Luce, J. and Kemmerer, D. (1996). Phonotactic and metrical influences on adult ratings of spoken nonsense words. In: *Proceedings of the Fourth International Conference on Spoken Language Processing*. Philadelphia. pp. 82-85, http://www.isca-speech.org/archive/icslp_1996/i96_0082.html, 11th November 2006.

Vitevitch, M., Luce, P., Charles-Luce, J. and Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40: 47–62.

Vitevitch, M. and Luce, P. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science* 9: 325–329.

Vitevitch, M. and Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40: 374-408.

Vitevitch, M and Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36: 481-487.

Vitevitch, M. and Luce, P. (2005) Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language* 52: 193-204.

Vitevitch, M., Luce, P., Pisoni, D. and Auer, E. (1999). Phonotactics, neighbourhood activation and lexical access for spoken words. *Brain and Language* 68: 306-311.

Warker, J. and G. Dell (2006) speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32 (2): 387–398.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science* 167: 392–393.

Warren, R. and Warren, R. (1977). Auditory illusions and confusions. *Scientific American* 223: 30-36.

Weber, A. (2000). The role of phonotactics in the segmentation of native and non-native continuous speech. In: *Spoken Word Access Processes*, A. Cutler, J. McQueen and R. Zondervan (eds.). Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. pp.143–146.

Weber, A. and Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America* 119: 597-607.

Webster's seventh new collegiate dictionary. (1967). Library Reproduction Service, Los Angeles.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence. *Applied Linguistics* 8(2): 95-110.

Wilson, M. (2003). Discovery listening - improving perceptual processing. *ELT Journal* 57 (4): 335-43.

Woodward, J. and Aslin, R. (1990). Segmentation cues in maternal speech to infants. A paper presented at the International Conference on Infant Studies, Montreal, Canada.

Yavaş, M. (2006). *Applied English Phonology*. Blackwell Publishing Ltd, Oxford.

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics* 18 (1): 49-68.

Appendix A.1 Consent and a survey filled in by non-native subjects prior to participating in the study.

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Researcher: Mr Faisal Aljasser Faisalmj24@hotmail.com

Dear Student,

You are asked to participate in this study because your English language level fits the researcher's criteria for study subjects participating in his PhD study. In this study you will be asked to take some tests both in the beginning and at the end of the term. Mr Aljasser will teach you the listening module in this term. However, this does not mean that participation is compulsory. Participation in this study is voluntary. You may choose to either take part or not to take part in the study. You may end your participation at any time by telling Mr Aljasser. Participation, non-participation, or ending your participation will not affect your grade in any way.

The information you provide and your test scores will remain confidential. If you agree to participate in this, please answer the questions below then sign the form.

1. How old are you? ....
2. What is your GPA scored at Level 2? ....
3. What is your Listening module mark scored at Level 2? ....
4. What is the average time a day (in minutes) you listen to native English speakers speaking whether through the media, internet or in the street? ....
5. Have you ever been taught rules regarding which English phonemes can and can not be adjacent in English syllables and words (e.g. /K/ can be followed by /w/ but /b/ can not? ....

My signature below indicates that I have read the information above and that I agree to participate in this study.

Name and signature of Participant ....

Date ....

Researcher's signature ....

Date ....

Appendix A.2 The thirteen non-words starting with consonant clusters that were used to judge which clusters are legal in Qassimi Arabic.

/dleʃ/, /mreʃ/, /bweʃ/, /bgeʃ/, /fweʃ, /dbeʃ/, /glauθ/, /flauθ//, /klud/, /grauθ//, /kweʃ/,

/θweʃ/, /breʃ/.

Appendix A.3 Transcription of the 6 non-words starting with illegal consonant clusters in both English and Qassimi Arabic with the sum of phoneme and biphone probabilities for the syllable starting at the second consonant of the non-word:

Nonword	**Sum of phoneme position probabilities**	**Sum of biphone probabilities**
/tlauθ/	.0512	.0008
/ʃlaudʒ/	.0455	.0003
/srud/	.1102	.0037
/hrɔɪz/	.0736	.0004
/nwʌz/	.0796	.0016
/mwʌθ/.	.0669	.0007
Mean	**.0711**	**.00125**

Appendix A.4 Transcription of the 6 non-words starting with illegal consonant clusters only in English with the sum of phoneme and biphone probabilities for the syllable starting at the second consonant of the non-word:

Nonword	**Sum of phoneme position probabilities**	**Sum of biphone probabilities**
/dlɔɪθ/	.0449	.0002
/dleʃ/	.071	.0024
/mreθ/	.0867	.0033
/mruð/	.0753	.0021
/bwaug/	.0479	.0001

/bwʌʃ/	.0672	.001
Mean	.0655	.00151

Appendix A.5 Statistics of non-words with high phonotactic probability:

Non-word	Initial phoneme position probability	Medial phoneme position probability	Final phoneme position probability	Sum of phoneme position probabilities	Initial-Medial biphone probability	Medial-final biphone probability	Sum of biphone probabilities
/tʃʌn/	.0089	.0392	.0961	**.1442**	.0005	.0057	**.0062**
/saɪv/	.1024	.0343	.0236	**.1603**	.0032	.0015	**.0047**
/sʌv/	.1024	.0392	.0236	**.1652**	.0059	.0012	**.0071**
/pɪv/	.0844	.0962	.0236	**.2042**	.0048	.0031	**.0079**
/pʌm/	.0844	.0392	.0494	**.173**	.0024	.0051	**.0075**
/saɪp/	.1024	.0343	.0371	**.1738**	.0032	.0014	**.0046**
	.0894	.0292	.0535	**.1671**	.0026	.0020	**.0046**
/pʌl/	.0844	.0392	.0737	**.1973**	.0024	.0046	**.007**
/pɪʒ/	.0844	.0962	.0017	**.1823**	.0048	.0003	**.0051**
/pæʒ/	.0844	.0794	.0017	**.1655**	.0087	.0002	**.0089**
/paɪt/	.0844	.0343	.0660	**.1847**	.0017	.0034	**.0051**
/pæv/	.0844	.0794	.0236	**.1874**	.0087	.0019	**.0106**
Mean				**.1754**			**.0066**

Appendix A.6 Statistics of non-words with low phonotactic probability

Nonword	Initial phoneme position probability	Medial phoneme position probability	Final phoneme position probability	Sum of phoneme position probabilities	Initial-Medial biphone probability	Medial-final biphone probability	Sum of biphone probabilities

/nɔɪv/	.0238	.0034	.0236	**.0508**	.0002	.0000	**.0002**
/θeʒ/	.0068	.0292	.0017	**.0377**	.0000	.0001	**.0001**
/wʌʒ/	.0203	.0392	.0017	**.0612**	.0006	.0000	**.0006**
/vaɪð/	.0224	.0343	.0031	**.0598**	.0025	.0002	**.0027**
/ʃaɪʒ/	.0097	.0343	.0017	**.0003**	.0000	.0457	**.0003**
/veʒ/	.0224	.0292	.0017	**.0533**	.0012	.0001	**.0013**
/ðʌv/	.0029	.0392	.0236	**.0657**	.0001	.0012	**.0013**
/ðaɪv/	.0029	.0343	.0236	**.0608**	.0001	.0015	**.0016**
/geʒ/	.0260	.0292	.0017	**.0569**	.0012	.0001	**.0013**
/vɔɪð/	.0224	.0034	.0031	**.0289**	.0004	.0000	**.0004**
/jʌʒ/	.0079	.0392	.0017	**.0488**	.0003	.0000	**.0003**
/ðɔɪʒ/	.0029	.0034	.0017	**.008**	.0000	.0000	**.0000**
Mean				**.0443**			**.0008**

Appendix A.7 20 Real-word fillers used in the Word Rating and Lexical Decision Tasks.

back, act, void, sack, soon, why, twist, fig, dry, boy, cheap, bus, test, joke, tow, red, car, out, find, cow.

Appendix A.8 Questionnaire used in the Word Rating Task.

You will listen to a list of 56 monosyllabic CCVC and CVC items. Some of these items are real English words and others are made-up words. When you decide that an item is a nonword, your task is to judge how English like it sounds by choosing a number from the scale below with number (1) representing least English-like and (5) representing most English-like. When the item is a real English word, you have to choose RW standing for Real Word.

Name :

---

(1) Very non-typical = 1 2 3 4 5 RW = Very typical
(2) Very non-typical = 1 2 3 4 5 RW = Very typical
(3) Very non-typical = 1 2 3 4 5 RW = Very typical
(4) Very non-typical = 1 2 3 4 5 RW = Very typical
(5) Very non-typical = 1 2 3 4 5 RW = Very typical
(6) Very non-typical = 1 2 3 4 5 RW = Very typical
(7) Very non-typical = 1 2 3 4 5 RW = Very typical
(8) Very non-typical = 1 2 3 4 5 RW = Very typical
(9) Very non-typical = 1 2 3 4 5 RW = Very typical
(10) Very non-typical = 1 2 3 4 5 RW = Very typical
(11) Very non-typical = 1 2 3 4 5 RW = Very typical
(12) Very non-typical = 1 2 3 4 5 RW = Very typical
(13) Very non-typical = 1 2 3 4 5 RW = Very typical
(14) Very non-typical = 1 2 3 4 5 RW = Very typical
(15) Very non-typical = 1 2 3 4 5 RW = Very typical
(16) Very non-typical = 1 2 3 4 5 RW = Very typical
(17) Very non-typical = 1 2 3 4 5 RW = Very typical
(18) Very non-typical = 1 2 3 4 5 RW = Very typical
(19) Very non-typical = 1 2 3 4 5 RW = Very typical
(20) Very non-typical = 1 2 3 4 5 RW = Very typical
(21) Very non-typical = 1 2 3 4 5 RW = Very typical
(22) Very non-typical = 1 2 3 4 5 RW = Very typical
(23) Very non-typical = 1 2 3 4 5 RW = Very typical
(24) Very non-typical = 1 2 3 4 5 RW = Very typical

(25) Very non-typical = 1 2 3 4 5 RW = Very typical

(26) Very non-typical = 1 2 3 4 5 RW = Very typical

(27) Very non-typical = 1 2 3 4 5 RW = Very typical

(28) Very non-typical = 1 2 3 4 5 RW = Very typical

(29) Very non-typical = 1 2 3 4 5 RW = Very typical

(30) Very non-typical = 1 2 3 4 5 RW = Very typical

(31) Very non-typical = 1 2 3 4 5 RW = Very typical

(32) Very non-typical = 1 2 3 4 5 RW = Very typical

(33) Very non-typical = 1 2 3 4 5 RW = Very typical

(34) Very non-typical = 1 2 3 4 5 RW = Very typical

(35) Very non-typical = 1 2 3 4 5 RW = Very typical

(36) Very non-typical = 1 2 3 4 5 RW = Very typical

(37) Very non-typical = 1 2 3 4 5 RW = Very typical

(38) Very non-typical = 1 2 3 4 5 RW = Very typical

(39) Very non-typical = 1 2 3 4 5 RW = Very typical

(40) Very non-typical = 1 2 3 4 5 RW = Very typical

(41) Very non-typical = 1 2 3 4 5 RW = Very typical

(42) Very non-typical = 1 2 3 4 5 RW = Very typical

(43) Very non-typical = 1 2 3 4 5 RW = Very typical

(44) Very non-typical = 1 2 3 4 5 RW = Very typical

(45) Very non-typical = 1 2 3 4 5 RW = Very typical

(46) Very non-typical = 1 2 3 4 5 RW = Very typical

(47) Very non-typical = 1 2 3 4 5 RW = Very typical

(48) Very non-typical = 1 2 3 4 5 RW = Very typical

(49) Very non-typical = 1 2 3 4 5 RW = Very typical

(50) Very non-typical = 1 2 3 4 5 RW = Very typical

(51) Very non-typical = 1 2 3 4 5 RW = Very typical

(52) Very non-typical = 1 2 3 4 5 RW = Very typical

(53) Very non-typical = 1 2 3 4 5 RW = Very typical

(54) Very non-typical = 1 2 3 4 5 RW = Very typical

(55) Very non-typical = 1 2 3 4 5 RW = Very typical

(56) Very non-typical = 1 2 3 4 5 RW = Very typical

Appendix A.9 A complete list of the items in the order they appeared in the Word Rating Task.

Practice items

/fɪk/ /ðɪθ/ /bwɔɪv/

Test items

1- /tlauθ/
2- /dlɔɪθ/
3- Back
4- /tʃʌn/
5- /saɪv/
6- Act
7- Void
8- /nɔɪv/
9- /θeʒ/
10- Sack
11- /ʃlaudʒ/
12- /dleʃ/
13- Soon
14- /sʌv/
15- /pɪv/
16- Why
17- twist
18- /wʌʒ/
19- fig
20- /vaið/
21- dry
22- /srud/
23- boy
24- /mreθ/
25- /mruð/
26- /pʌm/
27- cheap
28- /saɪp/
29- /ʃaɪʒ/
30- bus
31- /veʒ/
32- /hrɔɪz/

33- test
34- /bwaug/
35- joke
36- /pʌl/
37- /pek/
38- /ðʌv/
39- tow
40- /ðaɪv/
41- red
42- /nwʌz/
43- car
44- /bwʌʃ/
45- /pɪʒ/
46- /pæʒ/
47- /geʒ/
48- out
49- /vɔɪð/
50- /mwʌθ/.
51- /paɪt/
52- Find
53- /pæv/
54- /jʌʒ/
55- cow
56- /ðɔɪʒ/

Appendix A.10 Four additional real-word fillers used in the Lexical Decision Task.

Bridge, base, kiss, some.

Appendix A.11 three trial items used in the Lexical Decision Task.

/ki:v/, /kɔɪs/, table.

Appendix B. Items used in the word Spotting task

Appendix B.1 Practice items used before each list.

Sequence	Embedded word
θɑukhɑus	house
dɔɪkbuk	book
dɑuktʃeɪr	chair

Appendix B.2 10 real-word bearing fillers.

Sequence	Embedded word
sɔɪkhɔ:rs	horse
kɔɪskɪk	kick
sɔɪdspu:n	spoon
dɔɪsnaɪf	knife
ðaɪvmæn	man
ʃɔɪmkaɪnd	kind
mɔɪbnaɪt	night
ki:vsku:l	school
vɔɪkbaɪk	bike
dʒaɪztest	test

Appendix B.3 14 Bisyllabic nonsense sequences with no embedded words fillers.

tʃeftʌs - θæftʌdʒ - dɔɪftʌs - θeɪvnʌv - tɑuvnɔɪp - θɔɪvnʌp - zimθɔɪf - ʃɑumθeɪv - dʒɑumθɔɪt - tʃefhʌs - fɔɪmθʌs - pumθʌdʒ - θæfhʌdʒ - kɑufhʌp.

Appendix B.4 Embedded words with initial /l/.

Common Boundary	English Boundary	Qassimi Arabic boundary	No Boundary	Embedded Target Word
ʒi:tlɔ:rd	ʒi:dlɔ:rd	ʒi:blɔ:rd	ʒi:flɔ:rd	lord
fɑutlɔ:ŋ	fɑudlɔ:ŋ	fɑublɔ:ŋ	fɑuflɔ:ŋ	long
hɔɪtlɑ:rdʒ	hɔɪdlɑ:rdʒ	hɔɪblɑ:rdʒ	hɔɪflɑ:rdʒ	large
gɔɪtlɔ:st	gɔɪdlɔ:st	gɔɪblɔ:st	gɔɪflɔ:st	lost
zaɪtli:f	zaɪdli:f	zaɪbli:f	zaɪfli:f	leaf
vu:tleg	vu:dleg	vu:bleg	vu:kleg	leg
zɔɪtlenz	zɔɪdlenz	zɔɪblenz	zɔɪklenz	lens
fɔɪtlʌv	fɔɪdlʌv	fɔɪblʌv	fɔɪklʌv	love
zi:tlaɪk	zi:dlaɪk	zi:blaɪk	zi:klaɪk	like
vi:tlaɪn	vi:dlaɪn	vi:blaɪn	vi:flaɪn	line
zɑ:rtli:g	zɑ:rdli:g	zɑ:rbli:g	zɑ:rkli:g	league
zɑutluk	zɑudluk	zɑubluk	zɑukluk	look

Appendix B.5 Embedded words with initial /r/.

Common Boundary	English Boundary	Qassimi Arabic boundary	No Boundary	Embedded Target Word

ʒi:srʌsk	ʒi:mrʌsk	ʒi:ʃrʌsk	ʒi:grʌsk	rusk
fɑusraɪs	fɑumraɪs	fɑuʃraɪs	fɑugraɪs	rice
hɔɪsrʌst	hɔɪmrʌst	hɔɪʃrʌst	hɔɪgrʌst	rust
gɔɪsraɪt	gɔɪmraɪt	gɔɪʃraɪt	gɔɪgraɪt	right
zaɪsræʃ	zaɪmræʃ	zaɪʃræʃ	zaɪgræʃ	rash
zɔɪsraɪd	zɔɪmraɪd	zɔɪʃraɪd	zɔɪgraɪd	ride
fɔɪsrɪtʃ	fɔɪmrɪtʃ	fɔɪʃrɪtʃ	fɔɪgrɪtʃ	rich
zi:srɪsk	zi:mrɪsk	zi:ʃrɪsk	zi:grɪsk	risk
ʃɔɪsroul	ʃɔɪmroul	ʃɔɪʃroul	ʃɔɪgroul	role
zɑ:rsru:f	zɑ:rmru:f	zɑ:rʃru:f	zɑ:rgru:f	roof
mɔɪsrʌʃ	mɔɪmrʌʃ	mɔɪʃrʌʃ	mɔɪgrʌʃ	rush
zɑusraɪz	zɑumraɪz	zɑuʃraɪz	zɑugraɪz	rise

Appendix B.6 Embedded words with initial /w/.

Common Boundary	English Boundary	Qassimi Arabic boundary	No Boundary	Embedded Target Word
ðɑunweɪdʒ	ðɑubweɪdʒ	ðɑusweɪdʒ	ðɑukweɪdʒ	wage
sɔɪnweɪv	sɔɪbweɪv	sɔɪsweɪv	sɔɪkweɪv	wave
ðu:nwɔ:k	ðu:bwɔ:k	ðu:swɔ:k	ðu:kwɔ:k	walk
ðɔɪnweɪt	ðɔɪbweɪt	ðɔɪsweɪt	ðɔɪkweɪt	wait
dʒɑunweɪst	dʒɑubweɪst	dʒɑusweɪst	dʒɑukweɪst	waist
jɑunwi:k	jɑubwi:k	jɑuswi:k	jɑukwi:k	week
θu:nwæks	θu:bwæks	θu:swæks	θu:kwæks	wax
dʒaɪnwaɪt	dʒaɪbwaɪt	dʒaɪswaɪt	dʒaɪθwaɪt	white
zɑ:rnwaɪld	zɑ:rbwaɪld	zɑ:rswaɪld	zɑ:rθwaɪld	wild
ʃɑunwaɪf	ʃɑubwaɪf	ʃɑuswaɪf	ʃɑuθwaɪf	wife
nɔɪnwu:nd	nɔɪbwu:nd	nɔɪswu:nd	nɔɪθwu:nd	wound
zɑunwɪʃ	zɑubwɪʃ	zɑuswɪʃ	zɑuθwɪʃ	wish

Appendix C.1 The phonetic transcription task given to subjects one week prior to pre-tests.

Listening and speaking (Session 1)

Name: ______________________________

Write the phonetic transcription for each of the words below. If there are two pronunciations (i.e. British and American) for a single word, give the American one.

1- Lord / /
2- Long / /
3- Large / /
4- Lost / /
5- Leaf / /
6- Leg / /
7- Lens / /
8- Love / /
9- Like / /
10- Line / /
11- League / /
12- Look / /
13- Rusk / /
14- Rice / /
15- Rust / /
16- Right / /
17- Rash / /
18- Ride / /
19- Rich / /
20- Risk / /
21- Role / /
22- Roof / /
23- Rush / /
24- Rise / /
25- Wage / /
26- Wave / /
27- Walk / /
28- Wait / /
29- Waist / /
30- Week / /
31- Wax / /
32- White / /
33- Wild / /
34- Wife / /
35- Wound / /
36- Wish / /

Appendix C.2 A list of 12 illegal English consonant clusters given to subjects at the beginning of the treatment.

1- /dl/
2- /mr/
3- /bw/
4- /pm/
5- /bn/
6- /tf/

7- /dk/
8- /dg/
9- /kn/
10- /gm/
11- /fn/
12- /θb/