SECOND LANGUAGE PERCEPTION OF ACCENTED SPEECH

By
Ethan Andrew Cox

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree Of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2005
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3177525
Copyright 2005 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Ethan A. Cox entitled Second Language Perception Of Accented Speech and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Dr. Merrill Garrett

Dr. Janet Nicol

Dr. Natasha Warner

Dr. Ken Forster

Dr. James McQueen

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director:

Date: 12-22-04
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
ACKNOWLEDGEMENTS

This project could never have been completed without the support—academic and otherwise—of numerous institutions and individuals. I will attempt here to acknowledge all of them.

Foremost, I'd like to thank my committee; Merrill Garrett, Natasha Warner, James McQueen, Ken Forster and Janet Nicol. They all gave great advice and guidance on the implementation of these experiments and the interpretation of the results. Without their patience and willingness to advise over distance (and on short notice) I would never have finished.

I'd like to thank the Max Planck Institute and in particular Anne Cutler for making the Dutch aspects of this experiment possible. Also at the MPI, Rian and Ad facilitated administrative and technical aspects, respectively, of my stay.

At the University of Arizona, thanks go to the Center for English as a Second Language and in particular Dr. Kimberly, the Department of Spanish and Portuguese, and the Chicano/Latino Student Association. These organizations facilitated in subject recruitment and materials preparation.

Many thanks go to my colleagues during graduate school. In Arizona, Peter, Sheryl, Jason and Dave all found innumerable ways to keep me entertained and relatively sane throughout the years. They were aided and abetted by such luminaries as: Keith & Mary, Patrick, Ziya, Karen, and Lara. Connie was a great officemate and speech perception co-conspirator, and Dusana gave the office an international flavor. Becky & Robert befriended me from my arrival in Tucson as did Cathy who was additionally a great tenant. Steve & Missy and Roy both helped me to enjoy the great outdoors, while Alfonso and Wade were around for gaming, indoors.

At the Max Planck Institute I made many friends from The Netherlands and elsewhere who similarly eased my time in another country and culture. Thanks to Frank, Alissa, Greg, Anne Pier, Mirjam, Petra, Keren, Martijn and Fermin for many discussions and PhBeers. Thanks also to Michael and HenkJan for fun at the klimhal.

In the unique category of colleagues from both locales, I have to single out Lea and Dirk, two people who understand the importance of killing Dr. Lucky and of cloister clusters.

Finally, I joyfully acknowledge my dear friends Matt, Addy and mijn snoes, Jennifer. Your faith and encouragement, as well as you patience with my ups and downs, were instrumental and invaluable. My appreciation is boundless- Hartelijk bedankt!
DEDICATION

This dissertation is dedicated to my parents, Camille Cox, Babs Conant, Joe Cox and Bonnie Albert-Cox, and to the memory of Kerry P. Green.
TABLE OF CONTENTS - Continued

DISCUSSION ......................................................................................... 60
DISCUSSION: EXPERIMENTS 1A-1C .................................................. 60
CHAPTER 3: EXPERIMENTS IN ENGLISH AND SPANISH............ 72
EXPERIMENT 2A .................................................................................. 72
METHOD ............................................................................................. 73
Materials ............................................................................................. 73
Pretesting ......................................................................................... 73
Recording ......................................................................................... 74
Participants ..................................................................................... 75
Design ............................................................................................... 77
Procedure ....................................................................................... 77
RESULTS ............................................................................................ 78
DISCUSSION ....................................................................................... 80
EXPERIMENT 2B .................................................................................. 80
METHOD ............................................................................................. 81
Materials ............................................................................................. 81
Recording ......................................................................................... 81
Participants ..................................................................................... 82
Design ............................................................................................... 83
Procedure ....................................................................................... 83
RESULTS ............................................................................................ 84
DISCUSSION ....................................................................................... 86
EXPERIMENT 2C .................................................................................. 87
METHOD ............................................................................................. 87
Materials ............................................................................................. 87
Recording ......................................................................................... 87
Participants ..................................................................................... 87
Design ............................................................................................... 88
Procedure ....................................................................................... 88
RESULTS ............................................................................................ 88
DISCUSSION ....................................................................................... 90
DISCUSSION: EXPERIMENTS 2A–2C ................................................ 90
CHAPTER 4: GENERAL DISCUSSION................................................ 99
Summary Of Findings ......................................................................... 99
Signal Detection Analyses ................................................................. 101
Multiple Regression Analyses ......................................................... 104
Non-native Listener Analyses .......................................................... 105
Native Listener Analyses ................................................................. 106
Productive vs. Form of Stored Representations ................................ 108
FUTURE DIRECTIONS ......................................................................... 109
APPENDIX A: ENGLISH MATERIALS ............................................. 113
APPENDIX B: DUTCH MATERIALS ................................................ 115
TABLE OF CONTENTS - Continued

APPENDIX C: SPANISH MATERIALS ......................................................... 118
APPENDIX D: LANGUAGE BACKGROUND QUESTIONNAIRE AND
LEXICAL DECISION TESTS .............................................................. 121
REFERENCES .................................................................................. 123
LIST OF FIGURES

Figure 1. Mean subject reaction times for conditions in Experiment 1a. ..................47
Figure 2. Mean subject reaction times for conditions in Experiment 1b. ..................54
Figure 3. Mean subject reaction times for conditions in Experiment 1c. ..................59
Figure 4. Subject means for True responses in Experiments 1a and 1c. ..................64
Figure 5. Subject means for False responses in Experiments 1a and 1c. ..................65
Figure 6. Subject means for True responses in Experiments 1b and 1c. ..................69
Figure 7. Subject means for False responses in Experiments 1b and 1c. ..................70
Figure 8. Mean subject reaction times for conditions in Experiment 2a. ..................79
Figure 9. Mean subject reaction times for conditions in Experiment 2b. ..................85
Figure 10. Mean subject reaction times for conditions in Experiment 2c. ...............89
Figure 11. Subject means for True responses in Experiments 2a and 2c. ..................93
Figure 12. Subject means for False responses in Experiments 2a and 2c. ...............94
Figure 13. Subject means for True responses in Experiments 2b and 2c. ...............96
Figure 14. Subject means for False responses in Experiments 2b and 2c. ...............97
LIST OF TABLES

Table 1, Participant L2 Language Experience and Self-Report, Experiment 1a .......... 44
Table 2, Participant L2 Language Experience and Self-Report, Experiment 1b .......... 52
Table 3, Participant L2 Language Experience and Self-Report, Experiment 2a .......... 76
Table 4, Participant L2 Language Experience and Self-Report, Experiment 2b .......... 83
ABSTRACT

The present study addresses a core issue in the study of speech perception, the question of how stable phonological representations are accessed from an inherently variable speech signal. In particular, the research investigates the perception of accented English speech by native and non-native listeners. It is known from previous research that foreign-accented speech is harder for native listeners to process than native-accented speech. The reason for this lies in not only qualities of the input (deviation from native production norms, for example) but also in qualities of the listener. Specifically, listeners' speech perception systems are tuned from an early age to pay attention to useful distinctions in the language environment but to attenuate differences which are not useful. This quality of the listeners' speech processing system suggests that in addition to being native speakers of a language or languages, we are also native listeners.

However, what is a liability for native listeners (non-native input) may be a benefit for non-native listeners. When the foreign accent is derived from a single language shared between the speaker and the listener, application of native-language processing strategies to the accented input may result in more efficient processing of the input.

The experiments in this dissertation address this possibility. In an experiment involving Dutch listeners processing Dutch-accented and American English-accented sentence materials, a reaction time advantage was observed for the mutually-accented materials.
Experiments testing the main hypothesis with native Spanish-listening participants showed a different pattern of results. These participants, who had more experience with English overall than the Dutch participants, performed similarly to native-listening controls in displaying faster verification times for native accented materials than mutually-accented materials.

These experiments lead to the conclusion that native-like listening, as assessed by the sentence verification paradigm employed in these experiments, can be achieved by non-native listeners. In particular, non-native listeners with little experience processing spoken English benefit from hearing input produced in a matching accent. Non-native listeners with sufficiently more experience processing spoken English, however, perform similar to native listeners, displaying an advantage for native accented input.
CHAPTER 1: INTRODUCTION

Despite the apparent ease with which speech is processed and understood, the speech signal contains a great amount of variability from different sources. Different speakers, different speaking rates, ambient acoustics and background noise all contribute to this variability. How the listener is able to extract consistent linguistic representations from such a variable signal remains the core issue to be addressed by theories of speech perception.

Among the many sources of variability in speech, the variability imposed by a foreign accent constitutes perhaps the most perceptually salient but the least studied source. Foreign accent, which can be thought of as the imposition of a native phonology on a non-native language (Wingstedt & Schulman, 1987), is both ubiquitous and persistent in the production of adult learners of a second language. Even second-language learners with extensive exposure to their second language tend to produce speech with a foreign accent, though increased exposure is associated with reduced accentedness (Flege, Munro & Kay, 1995; Flege, Frieda & Nozawa, 1997). For a native listener, accented speech can be noticeably more difficult to understand than unaccented speech, both in terms of self report and performance (e.g. Munro & Derwing, 1995).

While accented speech can have deleterious effects on the comprehension of speech for a first-language (L1) listener, it could potentially pose a special problem for a second-language (L2) listener. On one hand, L2 listeners typically have considerably less experience processing their second language than their first, and thus could be hypothesized to face an especially difficult situation when confronted with accented L2
speech. Indeed, many studies have shown that even with extensive L2 processing experience, adult learners of a second language continue to process their second language 'through the ears of the first' (see, e.g. Cutler, 2002, for thorough review).

On the other hand, the persistence of L1 phonological processing strategies when listening to L2 speech allows the nonnative listener a potential advantage over the native listener when the accent in question is derived from a shared base language between the speaker and the listener. For example, unlike a native listener of English confronted with Dutch-accented speech, a native Dutch-listening interlocutor may find the non-canonical aspects of the speaker's accent to be predictable, since the misapplication of phonetic rules from the first language is quite possibly shared by the listener in his or her own productions. In this case, does the non-native listener have an advantage over the native listener?

The experiments undertaken in this dissertation were aimed at investigating this possibility. In a series of experiments involving perception of English by native listeners of Dutch (Experiment 1a) and Spanish (Experiment 2a), the potential benefit of a matching accent was explored using a sentence verification paradigm. Participants made speeded true/false decisions to sentences produced by a native speaker and a non-native speaker, the latter sharing an L1 background with the non-native participants. The prediction was that participants would more quickly and accurately process non-native accented speech than native accented speech when the accent in question was derived from a shared first language between the listener and the speaker.
Terminology

In this dissertation, the terms ‘native listener’ (NL) and ‘non-native listener’ (NNL) are used to describe the subject populations employed in this research. As experiments were conducted in three languages (English, Dutch and Spanish), a similar subject population may be NLs in one experiment but NNLs in another. Although these subjects were also native and non-native speakers, due to the focus of this research on perception and comprehension, whether participants spoke with or without an accent was tangential to the issue at hand; however, speaker–listener dyads from a shared language background are nonetheless at times designated as ‘mutually-accented.’ Where the speaking status of the participants is important (in stimuli recording, or describing results for example), ‘native speaker’ (NS) and ‘non-native speaker’ (NNS) will be so indicated.

The terms ‘first language’ (L1) and ‘second language’ (L2) are also used throughout the dissertation, to specify the native and non-native languages of the listeners and speakers. Specific characteristics of the participants (such as age of acquisition, amount of use, and self-rated proficiency) are given where appropriate. Note that all the L2 listeners in the study acquired L2 sometime beyond the age of four, resulting in a subject population ranging from early- to late-acquisition of L2 but not bilingual according to self report.

Three Areas of Relevant Research

The background and literature review is divided into three sections. The first section, *Perception of accented speech by native listeners*, is concerned with describing what is known about how native listeners process a specific type of variability in speech
perception, that of foreign accent. Although such difficulties are likely familiar to many readers, the experiments discussed in this section are aimed at identifying the nature and the extent of the difficulties encountered by native listeners when confronted with the distortions characteristic of a foreign accent. The next section, *Listening is language-specific*, presents the evidence that underlies the difficulty native listeners have in processing accented speech. Many listeners have first-hand awareness of the fact that different base languages give rise to qualitatively different accents, implying that experience interacts with speech production. However, the language-specificity of perception is less apparent consciously. The studies discussed in this section indicate that experience interacts with speech perception. The final section, *Perception of accented speech by non-native listeners* briefly summarizes the research which has shown that non-native listeners can show a benefit in processing their L2 when accented by their L1.

**Perception Of Accented Speech By Native Listeners**

The perceptual experience of listeners confronted with a foreign accent is often one of difficulty and reduced intelligibility. This experience was first quantified by Lane (1963) in an experiment studying the effects of different levels of background noise and different cutoff frequencies in broadband filtering on the perception of native and foreign-accented speech by native listeners. Using a word-identification measurement, Lane (1963) found graded adverse effects for both manipulations, added noise and filtering. Moreover, the accented speech was approximately 36% less intelligible than the unaccented speech across all conditions. These main effects, however, did not appear to interact, though no statistical tests were applied.
Though the research just cited provides empirical evidence that foreign accent can be more easily disrupted than native-accented speech, more direct evidence for processing difficulties as indicated by reaction time measurements was provided by Munro and Derwing (1995). They used a sentence verification task to investigate listeners’ processing of foreign-accented speech. The subjects were native listeners of English, and the sentence verification materials were English true/false statements produced by native-English accented and Mandarin Chinese-accented speakers. On each trial, subjects heard the stimulus three times, producing the reaction time measurement on the first presentation, and either an accent rating or a comprehensibility rating (counterbalanced) in the second and third passes. Results indicated that native listeners were significantly faster and more accurate for native- than foreign-accented speech. As predicted, Mandarin-accented sentences showed lower comprehensibility ratings and higher accent ratings than English-accented sentences. Further analysis revealed that tokens given a low comprehensibility rating took significantly longer to respond to than tokens rated moderate or high. Ratings of accentedness showed no effect. They additionally found that participants’ self-report of experience (defined as regular or occasional-to-no contact with Mandarin-accented speech), had no effect on any of their measures, including response times.

Bürki-Cohen, Miller and Eimas (2001) examined the impact of accented speech in a series of experiments using a measure of speech processing efficiency, the phoneme-monitoring task (PMT; Connine & Titone, 1996). Participants made phoneme monitoring responses to targets in words produced by a native and a non-native speaker
of English. The overall goal was to establish whether non-native speech led to lexically-influenced responses in the PMT. Targets were therefore placed word-initially in low- and high-frequency items so that the presence or absence of a statistically reliable frequency effect could serve as an index to indicate that phoneme monitoring responses were derived at least in part from stored, lexical representations of the target words. Bürki-Cohen et al. (2001) combined the native/non-native speaker manipulation with one other: the presence or absence of a secondary, lexically-driven task, specifically a concomitant noun/verb decision.¹

Experiments 1–8 examined the impact of accented speech when listening conditions were optimal—experimental tokens were presented in the clear with no background noise. Across these experiments, a consistent pattern of effects was observed. In all experiments with the secondary task, lexical frequency effects were obtained, in line with predictions and prior experiments (Eimas, Marcovitz Hornstein & Payton, 1990). However, frequency effects did not emerge in response to non-native stimuli. Instead, responses to non-native utterances patterned like responses to native utterances with frequency effects contingent on a secondary task.

Experiments 9–12 examined responses to native and non-native speech presented in multitalker babble noise; the presence or absence of a secondary task was again manipulated. As in Experiments 1–8, the secondary task resulted in significant frequency effects for responses to both native and non-native productions. However, the

¹ Bürki-Cohen et al. (2001) additionally manipulated the type of phonological feature contrasted between targets and nontargets. They found that decisions to voiceless targets were faster than decisions to voiced targets (exp 3,6,7,9,11) and decisions to alveolar targets were faster than decisions to labial targets (Exp 15 & 16). However, these factors never interacted with the other factors.
presence of background noise led to a different pattern of effects for the native and non-native stimuli. Responses to native productions were driven primarily by the perceptual information, as indicated by a lack of reliable frequency effects. Even at the highest SNR tested, -15dB, responses to native stimuli showed no frequency effects (Experiment 15). In contrast, responses to non-native stimuli yielded significant frequency effects at an SNR of only –3 dB (Experiment 11), indicating a shift in the locus of PMT responses for these stimuli. Thus, the processing burden imposed by presence of noise leads to increased reliance on lexical knowledge to make decisions about perceptual information for non-native stimuli.

Listening Is Language-Specific

Underlying native listener’s difficulties with accented speech are two distinct sources. The obvious source is the production values of the non-native speaker, and the distortions imposed by the non-native speaker’s base language. But a second source is the fact that listening to speech is, in part, a language-specific skill. Our speech perception capacities are optimized to process the ambient speech in our environment from an early age.

Infants begin developing their language listening skills before birth, as indicated by results showing that as early as two days after birth, infants prefer the native language of their mother over other languages (Moon, Panneton-Cooper & Fifer, 1993). However, since the input is filtered by the mother’s body, the speech information available to infants in utero lacks the fine phonetic detail necessary for robust language-specific development. Mehler and colleagues (e.g., Christophe, Mehler, & Sebastián-Gallés,
(2001); Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amieltison, 1988) have shown that infants display a preference for listening to their native language, and moreover that this preference extends to other languages with similar prosodic structure. For example, native English infants will reliably indicate perceiving a difference between English and French (which have dissimilar rhythmic structure) but not between English and Dutch (see Jusczyk, 1998 for a review.) At the level of prosodic processing, children have attuned to language-specific structural properties by the time of birth.

At a phonological level, infants are born prepared to acquire any of the world’s languages, that is, language-specific phonological development begins after birth. In foundational work, Eimas, Siqueland, Jusczyk and Vigorito (1971) examined young infants’ discrimination abilities for native and non-native phonological contrasts. They found that infants could discriminate both types of contrasts equally well. Werker and Tees (1984) showed that at 6 months, English-learning infants could readily discriminate a dental /t/ from a retroflex /ɾt/, a difference which is phonemic in Hindi but allophonic in English. Infants begin life exquisitely sensitive to information necessary to make many potential phonological contrasts.

This flexibility diminishes with increased exposure to their native language, however. The same infants who were capable of making non-native discriminations at 6 months were shown to have lost this ability by 10 months of age (Werker & Tees, 1984). Other studies by Werker and colleagues (Werker, Gilbert, Humphrey & Tees, 1981; Werker & Lalonde, 1988) attest to this diminishment in capability. From these studies, it appears that somewhere between 10 and 12 months of age, infants have identified the
contrasts required for successful acquisition of their native tongue and have lost their broader discriminatory abilities. It is also around this age that infants begin to show evidence of word-learning in both perception and production. Thus the sharpening of perceptual abilities may subserve (or be a product of) vocabulary acquisition itself.

With continued exposure, listening becomes more highly adapted to the perceptual and acoustical structure of the listener's native language. Experiments have demonstrated that adults use language specific phonological processing strategies optimized for native language input. For example, with respect to segmentation of continuous speech into words, Cutler and Butterfield's (1992) analysis of slips of the ear identified a significant tendency for English listeners to erroneously place syllable boundaries before strong but not weak syllables. A prior corpus study (Cutler & Carter, 1987) revealed that most English content words begin with strong syllables. Thus, in listeners' slips of the ear, they show evidence of exploiting a specific structural property of English when (mis)identifying word boundaries. Moreover, experimental evidence from word-spotting, such as Cutler and Norris (1988) have demonstrated English listeners' reliance on a segmentation strategy based on alternating strong- and weak-syllables.

As different languages have different structural properties, it is to be expected that native listeners of different languages will exploit the structural properties specific to their environmental language. In contrast to English, the rhythmic structure of French is syllable- rather than stress-based. Fittingly, French listeners show evidence of using a syllable-based rather than a stressed-based segmentation strategy during a syllable-
monitoring task (Mehler, Dommergues, Fraunfelder & Segui, 1981). In these experiments, native French listeners responded to targets that either matched the initial syllables of French words (pa- in pal.mier or pal- in pal.mier) or mismatched them. The mismatch was accomplished by reversing the target and the target-bearing word, i.e. monitoring for pa- in pal.mier and monitoring for pal- in pal.lace. Mehler et al. (1981) observed that these subjects responded fastest when the target aligned with the initial syllable, suggesting that syllables played an important role in continuous speech recognition.

However, English listeners presented with the same (French) materials as employed in Mehler et al (1981) did not show evidence of syllable-based segmentation (Cutler, Mehler, Norris, & Segui, 1986). Moreover, native French subjects in Cutler et al.'s (1986) experiment continued to employ a syllable-based segmentation strategy even when processing English words, which are comparatively difficult to reliably syllabify. Similar results have been shown for Spanish listeners (Sebastián-Gallés, Dupoux, Segui, & Mehler, 1992), another syllable-timed language.

The rhythmic structure of Japanese, in contrast to Spanish and English, is mora-based. Consequently, Japanese listeners show evidence of segmenting speech at moraic boundaries (Otake, Hatano, Cutler & Mehler, 1993; Cutler and Otake, 1994; but cf. Otake & Cutler, 1999 for evidence that sensitivity to a different language-specific feature—suprasegmental structure—may be acquired by non-native listeners.) Again, non-native listeners in these studies did not show the same sensitivity to the language-specific structural properties as native listeners did. Rather, the evidence from these
studies suggests that early and late acquirers of an L2 continue to apply their native-language segmentation strategies to processing non-native speech.

Interestingly, bilingual listeners show evidence of only possessing one rhythmical segmentation strategy. Cutler, Mehler, Norris and Segui (1992) examined bilingual French-English listeners’ performance in syllable-monitoring in both languages (using the materials of Cutler et al. 1986) and word-spotting in English (using the materials employed by Cutler and Norris, 1988). Overall performance showed a pattern unlike that of monolinguals in either language, and a group partitioning along the lines of father’s or mother’s native language failed to clarify the pattern of results. Likewise, partitioning by participant’s current country of origin did little to clarify the overall pattern of results. Remarkably, the bilinguals self-report of preferred language (as elicited by a justifiably famous query best paraphrased: “If you had to lose one or another language to live, which would you keep?”) effectively grouped the pattern of individual performance: participants who chose English patterned like monolingual English subjects (i.e., showing evidence of sensitivity to stress-based rather than syllable-based segmentation) while those who chose French employed syllable-based segmentation. Note, however, that the bilinguals in these experiments did not apply their singular segmentation strategy when it was inappropriate (i.e., syllable-based segmentation to English or stress-based segmentation to French), in contrast to the monolinguals in Cutler et al (1986) and Mehler et al. (1981).

The studies above explain why native listeners can have difficulty comprehending foreign-accented speech. When the listener is confronted with tokens that depart
radically from the canonical form, the native perceptual processing strategies will no longer optimally identify or categorize the input at the segmental level. In addition, the non-native speaker may be assimilating non-native contrasts to L1 norms, and in so doing produce contrasts the native-listener is not sensitive to. Fortunately, the listener in this case can call upon other processes—adaptation or recategorization at the perceptual level, and information at the lexical level—to aid comprehension. Is it possible, though, that there exists a situation in which non-canonical productions are easier, rather than harder, to process?

*Perception Of Accented Speech By Non-native Listeners*

Some evidence suggests that non-native listeners can process second-language speech more efficiently when it is accented by their native language than when it is accented by another base language. Eviatar, Leikin, Ibrahim and Sapir (2003) looked at fluent L2 Hebrew listeners, who were L1 listeners of Arabic and Russian. Using a gating paradigm, they showed that these subjects required more phonological information to accurately identify sentence-final words when they were produced in an unmatched foreign-accent (viz. American-accented Hebrew) than when they were produced in a matching foreign accent. However, for these subjects, no difference between a matched foreign accent and a native accent was observed. As these listeners were advanced L2 learners, it is possible that their non-native listening skills had developed to the point of being nearly fluent L2 listeners. On the other hand, the continued advantage in processing L2 when accented by their native language over other languages (as opposed
to difficulty with accented Hebrew regardless of the source language) is consistent with the interpretation that native-listening skills were still applied to some extent.

Experiments by Weber (2001a) underscore the persistence of native listening in non-native perception. Her experiments capitalized on a specific type of context-conditioned variability, phonological assimilation. Phonological assimilation occurs when a feature of a segment is altered as a consequence of coarticulation with either a preceding or subsequent segment (termed progressive and regressive assimilation, respectively). Exploiting the fact that this particular type of variation is language-specific, Weber (2001a) was able to adduce evidence for language-specific listening by native German listeners processing Dutch stimuli.

Specifically, Weber (2001a) asked German listeners to perform phoneme-monitoring in Dutch on monosyllabic items which were nonwords in both languages. The critical manipulation consisted of whether the nonwords contained sequences which, while legal in Dutch, violated obligatory assimilation rules in German. These listeners showed an inhibitory effect in monitoring for the target segments which violated German, but not Dutch, assimilation rules. These results were interpreted as evidence that German listeners, though processing Dutch, were still sensitive to the phonological structure of their native language. Native Dutch subjects performing the same task on the same materials showed no such effect, indicating that the source of the effect was in the listeners rather than in the materials themselves.

Weber’s (2001b) investigation considered also the role of phonotactic information in segmentation, using the word-spotting paradigm for both native and non-native
listeners. Phonotactics refers to the language-specific rules which dictate how phonemes may co-occur and in what environments. For example, in English, /ʃl/ is prohibited syllable-initially whereas /sl/ is allowed. In German, the opposite holds. This information could usefully help the listener to identify syllable boundaries, which in turn could be an aid to segmentation since every word boundary is necessarily a syllable boundary (though the opposite is clearly not so.) Extending the example above, an English listener hearing *Rush Limbaugh* need not consider the syllabification *Ru Shlimbaugh*, leading to more efficient lexical identification. Indeed, previous word-spotting experiments have established that Dutch listeners (van der Lugt, 1999) and English listeners (McQueen, 1998) exploit this property of their respective languages in segmenting continuous speech.

Weber (2001b) tested German and English listeners on English materials like those described above. Real word targets were embedded in larger non-word carriers in four conditions testing listener's sensitivity to the phonotactics of the two languages. Consider the environment of the English word *luck*, /lʌk/ in the token /pʌn lʌk/: the phonotactics of both English and German prohibit /n1/ as a syllable onset, forcing a syllable boundary clearly aligned with the embedded word. In the token /moi s1ʌk/, the phonotactics of German (but not English) require a syllable boundary between /s1/, while in the token /gɔi s1ʌk/ the restriction on /s1/ as a syllable-initial cluster is associated with English but not German. A final condition exemplified by the token...
/mɛɾfˈlʌk/ contained no restrictions on the unaligned syllabification, given that /f′l/ is licit syllable-initially in both German and English.

English monolingual listeners in these experiments indicated no sensitivity to German phonotactics but did display the expected sensitivity to English phonotactics. In contrast, German listeners who were highly proficient English users (they were professional translators) indicated a sensitivity to both German and English structural properties. This result underscores the persistence of native listening strategies when processing L2, though the results also indicate that sensitivity to a non-native structural property can be acquired to some degree.

Van Wijngaarden (2001) studied native (Dutch) and non-native (American) listener’s perception of Dutch produced by speakers of the same two language backgrounds. For native listeners, native speakers were found to be significantly more intelligible than non-native speakers, for all four speakers used in the study. However, for non-native listeners, the two most intelligible (i.e., least-accented) non-native speakers were found to be more intelligible than all four native speakers used in the experiment. Thus at least under some circumstances, non-native listeners can benefit from the characteristic distortions imposed on the speech signal by a non-native talker, when there is a mutual native-language background.

More direct evidence that non-native input can be beneficial for non-native listeners was seen in the results of Van Wijngaarden, Steeneken and Houtgast (2002). In this experiment, trilingual participants (L1 Dutch, L2 English and L3 German; ranking by age of acquisition, relative use and self-report of proficiency) were tested on native and
non-native productions of sentences in the same three languages. Proficiency was found to be a factor. When listening to English, subjects found the native speakers to be more intelligible than the non-native speakers. In contrast, when tested in German, non-native input was more intelligible than native input.

Similarly, in a recent study, Imai, Flege and Walley (2003) studied native and non-native listener’s word recognition for materials produced in a native and non-native accent, where the non-native listeners and speaker shared a first-language background. Imai et al. (2003) manipulated lexical variables in their materials set, specifically frequency and neighborhood density, in order to examine the contribution of post-phonological processing to word recognition. Materials were presented in background noise, as a means of avoiding ceiling effects for the native listeners. Their results indicated that native listeners performed better on native-accented stimuli than non-native accented stimuli, in line with predictions. Non-native listeners correctly identified more accented words than native-listeners. However, this effect was limited to words from high density lexical neighborhoods; the two listener groups performed equivalently for accented words from low-density neighborhoods. This interaction implied that processing at the phonological level can affect processing at the lexical level. In particular, when confusability at the lexical level is maximal (as in the case of words with many lexical neighbors), non-native listeners show a beneficial effect of hearing speech which is accented by their native language phonology.

Besides inefficiency at the phonological level of processing, non-native comprehension is further complicated by processing difficulties at the stage of lexical
selection. Some contemporary theories of spoken word recognition (e.g., Norris, McQueen & Cutler, 2000) posit that auditory lexical identification is the outcome of activation and competition between multiple representations consistent with the input. Thus hearing the word *speech* additionally activates the representations for *peach* and *each*, which are contained in the word, as well as to a lesser extent *speed* and *reach*, which overlap in the onset and rhyme, respectively, of the initial syllable. For the listener operating in L2, the possibility of interlingual activation and competition at this stage could lead to further processing inefficiency (Weber & Cutler, 2004).

A number of recent studies have used the eye-tracking paradigm to provide evidence that for bilinguals and advanced L2 listeners, the activation and selection process is not constrained by the language in use. For example, Spivey and Marion (1999) tracked Russian-English bilingual’s eye-movement in a task where participants were asked to move objects arranged before them, in Russian and in English. The objects included a target object (e.g. *marker*), two unrelated distractor items, and an interlingual distractor item whose name in Russian was phonologically similar to the target’s English name (e.g. *marku*; stamp). Even though participants were operating in English, fixation data indicated that the interlingual distractor attracted significantly more fixation time than the unrelated distractors, suggesting that lexical activation is language independent.

This language-independent lexical activation is sensitive to fine-grained acoustic-phonetic details in the input. Ju and Luce (2004) provided evidence of this in an eye-tracking experiment in which sub-phonetic manipulations were shown to impact the degree of language-independent activation at the lexical level. Specifically, native
Spanish listeners were presented with English objects in an eye-tracking paradigm similar to that of Spivey and Marion (1999), where participants were asked on each trial to select from among four objects displayed on a computer screen. The targets were Spanish words (e.g., playa, beach), presented with two unrelated distractors and an interlingual distractor whose name in English was phonologically similar to the target item's name (e.g., pliers). Importantly, the Spanish targets were manipulated such that half had Spanish-appropriate voice onset times and half had English-appropriate voice onset times. Results indicated that participants fixated on the interlingual distractor more than the control distractors when the Spanish target had the English-appropriate VOT then when it had the Spanish-appropriate VOT. However, fixations to the interlingual distractors when the targets had Spanish-appropriate VOTs were not significantly greater than fixations to the unrelated distractors, in contrast to the results of Spivey and Marion (1999).

The research of Bradlow and Bent (2002) was motivated by the observation that intelligibility is not only a function of the talker, but also of the listener. While in the case of foreign accent, it may be the case that the speaker's productions deviate from their canonical realizations, it is how a listener perceives these productions that plays the ultimate role. What may constitute a benefit for a particular speaker-listener dyad might be deleterious (or simply less beneficial) for another dyad.

Bradlow and Bent (2002) employed the clear speech register to investigate this speaker-listener relationship. Clear speech (Picheny, Durlach & Braida, 1985) is a mode speakers employ in situations where increased intelligibility is necessary, and has been
experimentally shown to lead to benefits for listeners with hearing difficulty (Schum, 1996), or when the environment is noisy (Payton, Uchanski & Braid, 1994). It was unknown, however, whether clear speech benefits would be observed for non-native listeners, as investigated by Bradlow and Bent (2002).

The benefits associated with clear speech observed for native listeners can be thought broadly to derive from two sources. Clear speech is, for example, characterized by having a reduced rate, more and longer pauses, and a wider pitch contour than conversational speech (Picheny et al., 1985). These modifications can be thought as increasing the salience of the signal itself, and could moreover be thought to hold when a clear speech mode is engaged in any language. Enhanced intelligibility as a result of such changes would presumably be observed for any cross-section of listeners, including non-native listeners.

Other acoustic consequences associated with clear speech seem, in contrast, to be language-specific. For example, clear speech is also characterized by stricter adherence to pronunciation norms, as shown in studies of coarticulation (Matthies, Perrier, Perkill & Zandipour, 2001). That is, speaking for increased intelligibility leads to more precision in articulation. Non-native listeners might reasonably be expected to benefit less (or not at all) from such code-enhancing consequences of clear speech articulation than native listeners.

Bradlow and Bent (2002) investigated native listeners and non-native listener's perception of English sentences produced in a clear speech and a conversational speech mode. Their measure of intelligibility (percentage of key words correctly identified in
transcription of auditorially-presented materials) indicated that native listeners showed a reliable benefit of approximately 15% for clear speech over conversational speech. In contrast, non-native listeners showed a significant benefit of only 5%, itself significantly smaller than that of the native listeners. This outcome was predicted by the logic outlined above: native listeners were able to take advantage of both the signal- and the code-enhancing properties of the clear speech, resulting in a large benefit to intelligibility. Non-native listeners, however, with less experience processing the language, benefited only from the broader, signal-enhancing qualities of clear speech, leading to a modest increase in intelligibility.

Bent and Bradlow (2003) provide further evidence that non-native listeners can benefit from input which is accented in their native language. In their experiments, native and non-native listeners participated in a sentence recognition task in which they heard sentences produced by native speakers and non-native speakers with a matched and a mismatched language background. Participants transcribed sentences presented auditorially, and keywords correct was the dependent measure. Their results indicated that native listeners performed best when processing native input; even the highest-proficiency non-native speakers produced less intelligible utterances than the native speakers for native listening participants. For non-native listeners, in contrast, the intelligibility of a high-proficiency non-native speaker from the same language background (i.e., 'mutually-accented') was found to be equal to the intelligibility of the native speaker. Interestingly, Bent and Bradlow (2003) also found that non-native listeners found non-native speakers from a mismatched L1 background to exceed the
intelligibility of native talker's productions when the non-native talker was highly proficient.

Bent and Bradlow (2003) discussed these results in terms of a 'matched interlanguage speech intelligibility' benefit and a 'mismatched interlanguage speech intelligibility' benefit, and argued for different sources of these two benefits. An 'interlanguage' (Selinker, 1972) describes the state of an L2 learner who has not yet mastered the phonology of the second language. While their productions in L2 are still influenced by their L1 phonology, they can also approach the canonical, native speaker norms for the language being acquired to a limited extent. In the case of the matched interlanguage benefit, Bent and Bradlow (2003) assumed that shared L1 phonological processes may be the source of the effect. In contrast, the mismatched interlanguage benefit may have been the result of similarities between the two base languages investigated—specifically Chinese and Korean. Alternately, the mismatched interlanguage benefit could arise from production strategies employed by all second language acquirers, such as hyperarticulation of contrasts which native speakers tend to hypoarticulate or fail to produce altogether.

Perhaps the most interesting evidence supporting the possibility that non-native listeners may benefit from speech accented by their native language comes from a case study reported by Eviatar, Leikin and Ibrahim (1999). The subject, R.K., was a native speaker of Russian with extensive knowledge of Hebrew who suffered a stroke, leaving her with transcortical sensory aphasia. Post-morbidly, the degree of impairment in processing both languages was generally equal, with deficits in her second language,
Hebrew, slightly more pronounced. Examination of patient R.K.'s phonological processing difficulties, as indexed by detection and discrimination tests, revealed an intriguing pattern. R.K. demonstrated great accuracy for Russian, and for Hebrew when spoken in a Russian accent. Accuracy dropped dramatically, however, for Hebrew when spoken in a native accent. Eviatar et al. (1999) interpreted this observation as consistent with the hypothesis that R.K.'s L2 phonology was processed via assimilation to her L1 phonological categories, and that it was this assimilation process itself that was damaged.

In summary, results of prior research supports the central hypothesis of the experiments conducted in this dissertation, that mutually-accented L2 input may incur less of a cost or potentially a benefit for non-native listeners while native listeners should process speech most efficiently when it is not accented.

The Current Experiments

The current studies employ the methodology of Munro and Derwing (1995) and extend the investigation to two novel subject groups, non-native listeners of English (L1: Dutch in Experiment 1a and L1: Spanish in Experiment 2a). Several modifications are made to the experimental design employed by Munro and Derwing (1995). One methodological adjustment, the use of single speakers for each accent and language factor, requires additional experiments which are themselves extensions of the Munro and Derwing (1995) experiments. In two cases, these experiments extend the manipulation of accent within the sentence verification paradigm to previously unexamined subject populations (native listeners of Dutch and Spanish; Experiment 1b and 2b, respectively) and in two cases examine processing in native English listeners on heretofore
unexamined foreign accents (Dutch- and Spanish-accented English in Experiment 1c and 2b, respectively). Some changes to the Munro and Derwing (1995) design, aimed at focusing the effects, are also incorporated.

First, as mentioned above, the current experiments employ the use of a single speaker rather than multiple speakers for each accent condition. On the one hand, the benefit of employing multiple speakers within each accent is in eliminating an explanation based on specific speaker qualities, allowing the accent to be composed of phonetic distortions common to all the non-native speakers—this avoids the confound of idiolect, or speaker-specific variation. On the other hand, use of multiple speakers also means that accent qualities are not necessarily consistent across speakers. While it is true that Munro and Derwing (1995) were able to informally classify the speakers chosen for recording as having moderate to heavy accents (the mean accent rating for the Mandarin speakers was 6.3 on a 9-point scale), they had no means of ensuring that individual speakers produced each sentence with the same features of foreign accent. For example, some speakers might have shown far more distortion in vowels than in consonants. With respect to specific words, some speakers might have been able to produce certain words closer to the native language norms than others—especially relevant if the truth value of a sentence is dependent on correctly identifying that specific word.

In the current study, therefore, single speakers are used for each accent condition. Since this conflates accent-specific and speaker-specific properties of speech, the same stimuli are used in a second experiment in which native English listeners respond to the same English-language stimuli used for the non-native listener experiment. If these
listeners show the expected effect for native listeners—faster responses to native accented than non-native accented speech—then the interpretation of the non-native listener experiment is clearer.

A second modification to the design of the current experiments (vis à vis Munro & Derwing 1995) address a potential confound of length and accent in their materials. Munro and Derwing (1995) reported a significant difference in the length of the materials produced by the native and the non-native talkers, indicating that native accented utterances were reliably shorter than Mandarin-accented utterances. Although measurement from sentence offset was intended to control for this overall difference in length, there remained a differing rate in the onset of information prior to participant’s decision points, which may have contributed to the advantage shown for native accented speech. This difference could lead to especially acute effects in message-level responses, due to the hypothesized cascade effect of difficulty in processing at multiple levels.

To prevent this, materials in the current experiment are equated for length between accents. The length of each token in each accent is measured and the two are averaged. Each token of the pair is then shortened or lengthened, via the PSOLA expansion/compression algorithm in Praat (Boersma, 2001) to the average length of the pair. In this way, the rate of speech information within levels of the Accent factor is reasonably equated.

The current study uses the sentence verification task (SVT), for two reasons. First, sentences provide the listener with more information than words. Not only does the use of sentences mean that both phonological and prosodic factors contribute to the input,
but sentences also provide more opportunities for phonological distortions to emerge. A second reason to use the SVT is that it focuses the participant’s attention on comprehension rather than the specific linguistic factors being studied. A number of recent studies (Pisoni, Manous & Dedina, 1987; Munro & Derwing, 1995; Clarke, 2003) have employed message-level measures paired with reaction times, and have shown them to be sensitive to phonetic-level manipulation.

The SVT additionally provides a measure of the contribution of higher-order strategy to the decision RT in the relative difference between responses to true and false propositions. Pisoni, Manous and Dedina (1987) employed this task in a series of experiments detailing the deleterious effects of synthetic speech compared to natural speech. Their experiments consistently found that natural speech was responded to faster than synthetic speech for both True and False items. Though statistical testing was performed separately on True and False responses, Pisoni et al. (1987) reported a numerical advantage for True sentences in their data.

Munro and Derwing (1995) did not obtain significant Truth Value effects in their experiments, though True responses were numerically faster than False responses. In addition, they did not obtain a significant interaction of the Truth Value factor with the Native Language of Speaker (Accent) factor, indicating that their accent effect was equally strong in both Truth Value conditions. Previous experiments using the SVT have, however, obtained significant truth value effects (e.g. Gough, 1966) Materials construction may have been responsible. Like the Pisoni et al. (1987) experiment,
Munro & Derwing (1995) did not create true/false minimal pairs in their sentence materials. Rather, true and false sentences varied freely.

The current experiments employ true/false sentence pairs which are minimally different, typically by one word. It is expected that this manipulation will lead to a more rigorous test of the Truth Value effect. While this effect is of less theoretical interest than the Accent effect, finding that non-native listeners are also able to capitalize on strategic guessing in a sentence verification task in their second language would be novel and provide a check on participant's L2 competence.

Overall, the prediction is that for second-language listeners, speech produced with an accent derived from a matched native language (e.g., Dutch listeners listening to Dutch-accented English) will be more easily and quickly processed than native-accented speech, as indexed by accuracy and reaction time measurements, respectively. Specific predictions for each experiment will be presented in Chapters 2 and 3.
CHAPTER 2: EXPERIMENTS IN ENGLISH AND DUTCH

Experiment 1a

In Experiment 1a, native speakers of Dutch made speeded True/False decisions to English sentences. Listeners heard sentences produced by a native speaker of American English and a native speaker of Dutch, resulting in native accented and non-native accented materials. No previous research has used the Sentence Verification Task to evaluate non-native listeners operating in L2, however, several previous studies have presented evidence that non-native listeners may benefit from mutually-accented speech (Van Wijngaarden, 2002a; Bent & Bradlow, 2003). Therefore, the prediction was that for Dutch listeners, the utterances produced by the non-native speaker would be easier and faster to process than the native-accented tokens, due to the listener’s and the speaker’s shared first language background.

METHOD

Materials

The experimental materials consisted of 40 English true/false sentence pairs, culled from a larger set generated by the experimenter. Each sentence consisted of a single clause statement, whose truth or falsity could be easily confirmed by real-world knowledge, e.g. (T) A pelican is a kind of bird; (F) A camel is a kind of bird. The sentences ranged from 3 to 8 words in length, and averaged 5.5 words and 7.49 syllables in length. The complete materials set is given in Appendix A.

Pretesting

Prior to recording, a larger materials set (consisting of 80 pairs of true/false sentences) was submitted to native speakers of Dutch for offline sentence verification in
English, to establish that the selected materials could be accurately verified by second-language listeners, specifically with respect to vocabulary. Ten native speakers of Dutch, with similar language background characteristics to the participants in the main experiment (see below) served as participants for this task. The final, 40-item set was comprised of all pairs that had been responded to correctly by 8 or more participants, with the exception of two sentences which, along with their true or false counterparts, were set aside as practice items. For these items, either the true or the false version of the sentence pair had been correctly verified by only 7 of the 10 participants. Two further sentence pairs (containing proper nouns and thus problematic for a later, Dutch-language condition; see Recording in Experiment 1b, below) were also used as practice items. The final materials set thus comprised 36 experimental sentence pairs and 4 practice sentence pairs.

Recording

The experiment required two speakers: A native speaker of Dutch proficient in English, and a native speaker of American English proficient in Dutch (Dutch proficiency was required for recording stimuli in Dutch for a necessary control experiment, Exp. 1b, below). Native speakers of Dutch were recruited on the University of Arizona campus and in the Tucson community. Recording a native Dutch speaker in the United States, rather than in The Netherlands, was preferred for two reasons. First, it was essential to the control condition that the selected speaker had American English as a target pronunciation; in The Netherlands, students acquiring English are generally taught British pronunciation norms. Even students studying American English specifically are
likely to encounter more British-accented speech outside of their studies, given geographical factors and the wide availability of British media. This situation could have led to an accent confound, whereby a combination of Dutch and British English accents would be present. Second, a speaker with a moderate, rather than a heavy accent, was the aim. Speakers who had spent some time living in an American English environment were thus preferred.

A total of 5 native speakers, with varying language backgrounds with respect to acquisition and usage of English and length of residency in the United States, were recorded uttering the materials set in English. Based on informal feedback from native speakers of Dutch within the Max Planck Institute for Psycholinguistics, Nijmegen (henceforth MPI Nijmegen) community, one of these, BN, was selected for the experimental sentences. BN was a native speaker of Dutch from the same region as the projected listener population (Brabant, The Netherlands). As a professor in the US for approximately 14 years and with an American spouse, BN’s English experience was extensive, though BN reported using Dutch frequently, in reading Dutch media and in contact with relatives and friends. Practice tokens were taken from another speaker’s recordings.

Native speakers of American English with sufficient proficiency in Dutch proved difficult to find in the University of Arizona and Tucson community. Recordings were solicited of several individuals across the US and of one speaker from the University of Arizona. Advice from native speakers of Dutch within the MPI Nijmegen community aided the final selection. Speaker NW was selected based on recording quality and
accentedness in Dutch production, the latter again for reasons related to Experiment 1b. This speaker’s exposure to and proficiency in Dutch was considerably less than the NNS BN’s familiarity with English. Productions by another speaker were used for practice tokens, as above.

Both speakers, BN (L1: Dutch) and NW (L1: American English) were recorded at the University of Arizona. Recording of the (English) materials for this experiment and the (Dutch) materials for Experiment 1b was conducted at the same time. Recording sessions were conducted in a sound-attenuated booth using a high-quality microphone (ElectroVoice RE27N/P) and a digital recording device (Alesis ML-9600). Each speaker was recorded in a single session. Recording was blocked by language, and speakers were recorded in their native language first. Speakers were given the list of recording materials immediately prior to the recording session. They were given some time to go over the list and to ask the experimenter how to pronounce any unfamiliar words. Prior to recording, the speakers also went through the list once aloud, with the experimenter monitoring for and correcting any gross mispronunciations. During the recording, each sentence was uttered three times at a normal speaking rate. The experimenter monitored the ongoing recording and asked the speaker to repeat any sentences uttered with obvious disfluencies or pronunciation errors.

Sentences were subsequently transferred to a computer for editing and experimental presentation. At this stage, particular tokens were selected from among the repetitions of each sentences. First, tokens were selected from the NNS productions. The procedure was to choose the first of the three utterances which was acceptable with
respect to production and recording quality. The choice of the NS tokens was not constrained by questions of production quality. Rather, NS token choice was driven by overall length—the token which was closest in length to the matching NNS token was chosen. Durations of both selections were noted for later analysis, and all tokens were equalized for volume by means of RMS-equalization, using Cool Edit software.

The selected sets of stimulus items for each speaker were digitally manipulated in one further way. Each specific True or False token had a native speaker and a nonnative speaker recording. These differed in length, thus introducing a possible confound of length and accent/speaker (the accent/speaker confound was considered in Experiment 1c). To address this, the PSOLA expansion/compression algorithm within the Praat program (Boersma, 2001) was used to modify each token pair to the averaged length of the pair for all differences greater than 1% (all but 10 pairs). Before compression/expansion, the NS tokens averaged 1516 ms in length, and the NNS tokens averaged 1608 ms in length, a difference of 106 ms. Thus, for most token pairs, NS tokens were expanded, (mean 3.36%, range −7 to 11 %) and NNS tokens were compressed (mean 96.64%, range 88% to 106%) Post-manipulation, tokens were judged to sound natural.

Participants

Twenty-four native speakers of Dutch were recruited from the MPI Nijmegen subject pool to participate in this experiment, and were compensated with a small amount of money for their time. Most participants were students at the nearby University of Nijmegen, and in the typical college-age range (18–22 years). All were proficient in
English and had similar linguistic backgrounds with respect to acquisition and usage of 
English. Only participants who claimed to be bilingual were excluded from the 
experiment, however no participants met this criterion. Most participants were familiar 
with a 3rd or 4th language, these most frequently being French and German, though only 
4 participants reported using either language more than English. Participants’ experience 
with English was evaluated after behavioral testing by means of a questionnaire (which 
queried such factors as age of acquisition, amount of daily use of English, and subjective 
proficiency; See Appendix D), and an unspeeded lexical-decision test intended to index 
participants’ English lexical knowledge. None of words in the lexical decision test 
appeared in the experiment. A summary of these data is presented in Table 1.

Table 1: Participant L2 Language Experience and Self-Report, Experiment 1a

<table>
<thead>
<tr>
<th>Measure</th>
<th>Age</th>
<th>AOA</th>
<th>YOE</th>
<th>Self-report English Experience</th>
<th>LexDec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R/exp</td>
<td>W/exp</td>
</tr>
<tr>
<td>Mean</td>
<td>20.9</td>
<td>10.7</td>
<td>9.9</td>
<td>4.9</td>
<td>3.8</td>
</tr>
<tr>
<td>Minimum</td>
<td>18.0</td>
<td>6.0</td>
<td>6.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>26.0</td>
<td>14.0</td>
<td>20.0</td>
<td>7.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

AOA = age of acquisition; YOE = years of experience; R/exp = reading experience; W/exp = writing experience; S/exp = speaking experience; LexDec = percent correct on offline lexical decision test.

Design

The overall design was Accent (Act: NS, NNS) x Truth Value (Tval: True, False), 
both factors varying within-subjects. It was important that participants not hear both the
true and false versions of a sentence pair, nor the same item spoken by both the NS and
NNS, as this would have led to repetition or practice effects, resulting in speeded
decisions. To prevent this, 4 lists were used for item counterbalancing, between subjects.
Six participants heard each list. Each list was pseudorandomized to ensure that no more
than 3 consecutive instances of a particular stimulus factor (T/F, NS-/NNS-accented)
would be presented.

Procedure

Subjects were tested at the MPI Nijmegen in The Netherlands. Subjects were
tested individually or in groups of up to three people in a quiet room at individual testing
stations divided by baffles. Each station provided a monitor, high-quality headphones,
and a labeled response button box. Presentation of experimental materials and reaction-
time measurement was controlled by a computer running the NESU experimental
software developed by the MPI Nijmegen. Sentences were presented over the
headphones at a comfortable listening level. Participants were given instructions on-
screen, in the language of the Sentence Verification Task (English). Subjects were
instructed to make their verification decisions as quickly and accurately as possible.
Assignment of the True and False RT keys was counterbalanced across handedness of
participants; half had True assigned to their dominant hand. Participants underwent a
short practice session (4 trials) prior to the presentation of the 36 experimental trials, in
order to familiarize them with the trial structure. SOA and timeout was 5000
milliseconds. After behavioral testing, participants responded to the language
background questionnaire and the offline lexical decision test. The session took about 20
minutes, about 10 minutes for the experiment and 10 minutes for the pencil-and-paper tests.

RESULTS

Recorded RTs were first adjusted to reflect measurement from the offset of the sentence. Responses were evaluated for accuracy. RTs associated with incorrect responses were dropped from analysis. Outliers, defined as reaction times falling outside the range of –250 ms to 2500 ms, were also excluded from analysis. A negative lower cutoff was deemed necessary since much research has indicated that words, especially in context, are often identified before their acoustic offset (Marslen-Wilson, 1987; but cf. Bard, Shillcock & Altman, 1988). The upper limit, 2500 ms, was set relatively high to allow for the fact that participants were operating in their second language, and might reasonably be expected to take somewhat longer than usual. Combined, outliers and errors accounted for 6.4% of the data. No individual subjects (max error: 22.22%, 1 Subject) or items (max error: 25%, 1 Item) exceeded 25% errors, so no subjects or items were dropped from analysis. Figure 1 shows the mean reaction times for the critical conditions in Experiment 1a.
These RT data were then subjected to two ANOVAs. The first considered subjects as the random factor \(F_1\) and consisted of a \(2 \times 2 \times 4\) design which crossed the factors of Truth Value(2) and Accent(2) within-subjects and List(4) between subjects. The second analysis treated Items as the random factor \(F_2\) and crossed the factor of Truth Value(2) between items\(^2\) with the factor of Accent(2) within items. This yielded a significant main effect of Truth Value in both analyses, \((F_1(1, 20) = 91.08, p < .01\) and \(F_2(1, 70) = 10.99, p < .01\).\) The main effect of Accent was significant in both analyses.

---

\(^2\) It should be noted that, since the items were minimally different with respect to the true/false manipulation, in analysis by items, truth value could have been treated as a within-items variable. However, on the logic that the truth value manipulation was of a different character than the accent manipulation, since no actual token could be both true and false, this factor was treated as a between-items manipulation in statistical analysis.
The interaction of Truth Value and Accent showed a trend towards significance in analysis by subjects, \((F_{1}(1,20) = 3.91, p = .06)\), but not in analysis by items, \((F_{2}(1,70) = 1.47, p = .22)\). No other main effects or interactions were significant in either analysis.

ANOVA was also applied to error rates, using the same models as outlined for the RT analysis. Error rates were low overall, and no significant main effects or interactions were observed. The mean error rates for native English speech were 6.9% and 5.1% for True and False, respectively. The error rates for non-native English speech were 7.4% and 6.0% for True and False, respectively.

DISCUSSION

The finding of an overall effect of Accent indicates clearly that for these listeners, Dutch-accented English was easier to process than native-accented English, as predicted. Moreover, it is clear that numerically, this advantage is stronger for false responses than for true responses; the advantage observed for false responses is about 148 ms, while the advantage observed for true responses is about 50 ms, though the Truth Value x Accent interaction indicates only a trend towards statistical significance in analysis by subjects. A ready explanation can be found in considering listener’s guessing strategies. As information unfolded over time, participants were possibly pre-activating various candidates for the final word of the sentence, upon which the sentence’s truth or falsity turned. When the sentence ended as predicted, RTs were low and the difference in processing speed between the accented and unaccented tokens was thus minimized. As a result of this strategy, False sentences became a more reliable test bed for the impact of
mutually-accented speech. When participants guesses for the sentence-final word were supported by the input, the subjects were forced to process the final word in greater detail, and discrepancies between the stored representation and the utterance resulted in longer RTs.

Experiment 1b

In Experiment 1b, native listeners of Dutch made speeded True/False decisions to sentences presented in Dutch. Importantly, listeners heard sentences produced by a native speaker of Dutch and a native speaker of American English, resulting in native accented and non-native accented materials. The same speakers who produced English stimuli in Experiment 1a were used, now speaking Dutch. This resulted in a reversal of the speaker associated with native and non-native speech. In Dutch, speaker BN produced native accented speech while speaker NW produced non-native accented speech. The prediction, consistent with similar experiments involving the perception of accented speech (e.g. Munro & Derwing, 1995) was that the utterances produced by the native speaker would be easier and faster to process than the foreign-accented tokens.

METHOD

Materials

The experimental materials consisted of 40 Dutch true/false sentence pairs, generated by the experimenter. These sentences were the translation equivalents of the materials used in Experiment 1a. The translations were generated by the experimenter, with input from native speakers of Dutch. As in Experiment 1a, each sentence consisted of a single clause statement, whose truth or falsity could be easily confirmed by real-
world knowledge, e.g. (T) *Een pelicaan is een soort vogel (A pelican is a type of bird)*; (F) *Een kameel is een soort vogel (A camel is a type of bird)*. The completer materials set can be seen in Appendix B. Sentences ranged from 3 to 9 words in length, and averaged 5.5 words and 8.38 syllables in length.

**Pretesting**

Prior to recording, the materials set was submitted to native speakers of Dutch for offline sentence verification in Dutch, to establish the accuracy of the translations as well as to verify the cross-linguistic and cultural validity of the statements. Ten native speakers of Dutch, with similar language background characteristics to the participants in the main experiment (see below) provided these data. All items were responded to correctly by 9 or more participants, with the exception of two sentences which had already been established as practice sentences in the previous experiment.

**Recording**

Stimulus recording procedures were identical and contemporaneous with the procedures outlined in Experiment 1a. The same two speakers were used as in Experiment 1a, with a reversal in the speaker associated with native accented and non-native accented speech. Specifically, speaker BN (L1: Dutch) produced native accented Dutch while speaker NW (L1: English) produced non-native accented Dutch. Informal judgments by native listeners of Dutch indicated that speaker BN produced Dutch with a native accent while speaker NW produced Dutch with a moderate and identifiable American English accent.
As with the English stimuli used in Experiment 1a, the Dutch items were equated for length using the PSOLA expansion/compression algorithm in Praat. Eleven pairs had length differences of less than 1% and were therefore unchanged. Of the remaining items, the Dutch items on average were produced more quickly by the NS than the NNS (mean length 1677 ms and 1815 ms, respectively; a difference of 137 ms). This pattern was consistent with that observed for the materials in Experiment 1a, that is, the NNS was slower than the NS. As a consequence, NS utterances were lengthened 4.14% on average (range –5% to 12%), NNS utterances were shortened on average to 96.37% of their original length (range 90% to 105%). The final average length of the tokens was 1746 ms.

Participants

Twenty-four native speakers of Dutch were recruited for this experiment. As with the subject group in Experiment 1a, these volunteers were recruited through the MPI Nijmegen subject pool and consisted of college-age individuals primarily recruited at the nearby University of Nijmegen. Participants were paid for their participation. Although these participants were listening in their native language, Dutch, proficiency and background information on English were collected as in Experiment 1a for the purpose of ensuring parity between subject groups. These participants also rated the degree of foreign accent of the NNS using a 7-point Likert scale for which 1 indicated a very light or no accent and 7 a very heavy accent, and were asked to guess the native language of the foreign-accented speaker. No bilingual participants were included. Though most subjects were familiar with a 3rd or even 4th language (most frequently German and
French) only a few participants (N = 5) claimed to use this language more frequently than English. None of the participants had lived in an English-speaking country for an extended time. Details of these data are summarized in Table 2.

Table 2: Participant L2 Language Experience and Self-Report, Experiment 1b

<table>
<thead>
<tr>
<th>Measure</th>
<th>Age</th>
<th>AOA</th>
<th>YOE</th>
<th>R/exp</th>
<th>W/exp</th>
<th>S/exp</th>
<th>LexDec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>20.0</td>
<td>11.7</td>
<td>8.3</td>
<td>5.0</td>
<td>3.3</td>
<td>3.5</td>
<td>49.7%</td>
</tr>
<tr>
<td>Minimum</td>
<td>18.0</td>
<td>10.0</td>
<td>5.0</td>
<td>3.0</td>
<td>1.0</td>
<td>1.0</td>
<td>15.0%</td>
</tr>
<tr>
<td>Maximum</td>
<td>24.0</td>
<td>14.0</td>
<td>13.0</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>75.0%</td>
</tr>
</tbody>
</table>

AOA = age of acquisition; YOE = years of experience; R/exp = reading experience; W/exp = writing experience; S/exp = speaking experience; LexDec = percent correct on offline lexical decision test.

**Design**

The experimental design of Experiment 1b was identical to that of Experiment 1a. The factors of Truth Value and Accent were fully crossed within subjects but between items, yielding a 4-list counterbalancing scheme. Pseudorandomization parameters were identical to Experiment 1a. Each list was presented to 6 participants.

**Procedure**

Experimental procedures for Experiment 1b were identical to those of Experiment 1a, with a few exceptions. Subjects were again run at the MPI Nijmegen. Subjects were run individually and in groups of up to three in the same room and using the same apparatus as in Experiment 1a. In this experiment, subjects were given instructions in
Dutch, rather than in English; instructions were again presented on-screen. The instructions were translated from the instructions used in Experiment 1a by the experimenter with input from native speakers.

RESULTS

Recorded RTs were first adjusted to reflect measurement from the offset of the sentence. Responses were evaluated for correctness; RTs associated with incorrect responses were dropped from the analysis. Scores less than -250 ms and greater than 2500 ms were deemed outliers and also dropped from the data set. Subject error rates averaged 5.56%, maximum score 13.89%; Item error rates indicated one item (No. 16; see Appendix B) had an error rate exceeding 25% and this item was thus dropped from further analysis (final Subject N = 24; Item N = 35).

RT data were subjected to an ANOVA with subjects as the random factor (F1), as well as an ANOVA with items as the random factor (F2). With respect to statistical design, these analyses were identical to those of Experiment 1a. These analyses yielded a significant main effect of Truth Value by subjects and by items, (F1(1, 20) = 28.31, p < .01; F2(1, 68) = 4.28, p < .05.) The Accent factor showed robust effects in both analyses, (F1(1,20) = 71.75, p < .01; F2(1,68) = 18.79, p < .01), as did the Truth Value by Accent interaction, (F1(1,20) = 40.85, p < .01; F2(1,68) = 7.17, p < .01.) Analysis by subjects indicated two further higher-order effects, the interaction of Accent x List, (F1(1,20) = 3.95, p < .05), and the three-way interaction of all variables, (F1(1,20) = 14.03, p < .01), both of which may be considered circumstances of subject variation.
The means for the four critical conditions can be seen in Figure 2. Planned comparisons of the Accent effect for each level of the Truth Value factor indicated that the 210 ms difference between the NS and the NNS within False responses was significant, \( t(23) = 8.51, p < .01 \), but that the 41 ms difference observed between the accent conditions within True responses was not significant, \( t(23) = 1.32, p = .19 \).

Experiment 1b: Dutch Listeners

![Figure 2](image_url)  
*Figure 2. Mean subject reaction times for conditions in Experiment 1b. Subjects were native Dutch listeners, sentences were in Dutch. Error bars represent standard error values.*

Error rates were also analyzed using the same models as for the RT analysis. These analyses yielded no significant effects, owing to low error rates. Native Dutch stimuli showed 4.2% and 5.3% error rates in True and False, respectively. Non-native Dutch stimuli received 5.3% errors in both Truth Value conditions.
With respect to participants’ accent ratings for the non-native speaker, the mean rating on the 1-7 scale (where 1 indicated a light or no accent) was 4.5 (s.d. .2). Many participants accurately identified the native language of the NNS as English, (N = 12; 50%) generally, though somewhat fewer specified American English precisely (N = 10; 41.6%).

**DISCUSSION**

These results indicate a robust beneficial effect of a native accent for listeners processing their native language. When Dutch listeners were presented with native and non-native productions, a clear advantage for the native accented productions was observed. These results are consistent with prior experiments showing a reaction time advantage for native accented speech over non-native accented speech using the SVT (Munro & Derwing, 1995). Moreover, these results compliment the findings of Experiment 1a, showing that the effects observed in that experiment were not due solely to the fact that listeners were operating in their non-native language. In both experiments, reaction times associated with the native speaker of Dutch were faster than those associated with the native speaker of English, regardless of the language of the stimuli.

Also consistent with Experiment 1a was the observation of stronger accent effects for False responses then for True responses, indicated by the significant interaction of the two factors and the outcome of planned comparisons. Presumably, higher-order strategic guessing factors influenced True responses more so than False responses, leading to a stronger phonological influence in processing the False responses.
Note however, that between the two experiments, faster RTs are associated with the same speaker, allowing for an explanation based on specific speaker characteristics other than (or in addition to) accent quality. For example, an interpretation of greater clarity in speech by the native Dutch speaker than the native English speaker, regardless of which language they were using would also account for these results. In order to address this line of argument, Experiment 1c was performed.

**EXPERIMENT 1C**

In Experiment 1c, native listeners of American English made speeded true/false decisions to English sentences. Specifically, listeners heard sentences produced by a native speaker of English and a native speaker of Dutch, thus constituting native and non-native accented materials. These were the same materials as used in Experiment 1a. The simple prediction for these English listeners, was that the utterances produced by the native speaker would be easier and faster to process than the non-native accented tokens. Importantly, this outcome, coupled with the results of Experiment 1b, would eliminate an explanation of the results of Experiment 1a appealing to specific speaker characteristics other than accent, such as overall clarity in production.

**METHOD**

*Materials*

The experimental materials were those of Experiment 1a, 36 English true/false sentence pairs and 4 practice true/false sentence pairs. The same audio files were employed as were used in Experiment 1a.
Pretesting

Pretesting of the materials was undertaken in Experiment 1a and 1b. No further pre-testing was required for Experiment 1c.

Recording

Recording procedures were as outlined in Experiment 1a.

Participants

Twenty-four native speakers of American English were recruited from the University of Arizona Psychology department subject pool. Participants were paid a small amount of money for their time. Most were students, in the typical college-age range (18–22). Participants were assessed post-experimentally for familiarity with Dutch and Dutch-accented English in particular, but as none were fluent or familiar with Dutch, none were replaced. In addition, participants were queried on their experience with second languages more generally, but none reported more than basic fluency in a second language, most often Spanish.

Design

The experimental design was identical to that of Experiments 1a and 1b. The factors of Truth Value (Tval: True, False) and Accent (Act: NS, NNS) were fully crossed within subjects, with item counterbalancing across four lists. The lists were pseudorandomized to ensure no more than three consecutive instances of a particular factor occurred. Each list was presented to 6 participants.
**Procedure**

The experimental procedure was similar to that of Experiments 1a and 1b, with the following differences: Subjects were tested at the University of Arizona, rather than at the MPI Nijmegen. Subjects were tested individually in sound-attenuated booths rather than in groups. Experimental presentation and reaction time measurement was controlled by a computer running DMDX (Forster & Forster, 2003) software, rather than NESU. Finally, these participants received a modified language background questionnaire to probe for L2 experience and experience with Dutch in particular. Participants were not given the lexical decision test (as they were native listeners). They were also asked to generate accent ratings for the non-native speaker and were asked to guess the origin of the NNS's accent. As in Experiment 1b, the instrument was a 7-point Likert scale for which 1 indicated a very light or no accent and 7 a very heavy accent. All other procedural details were identical to those of the previous experiments.

**RESULTS**

Recorded reaction times were first adjusted to reflect measurement from the offset of the sentence. RTs associated with errors were dropped. Scores outside the ~250 to 2500 ms range were similarly treated as errors, and evaluation of subject and item performance was conducted on this reduced data set. No subjects exceeded 16.67% errors, however one item pair (No. 5; see Appendix A) was dropped from further analysis based on a high error rate, 29%. The final subject mean percent error was 3.94%. The RT means for the 4 critical conditions are presented in Figure 3, below.
As with Experiments 1a and 1b above, by-subjects and by-items ANOVA was applied to the RT data, according to the identical statistical designs for $F_1$ and $F_2$. These analyses yielded a robust main effect of Truth Value, ($F_1(1,20) = 24.86, p < .01$; $F_2(1,68) = 9.38, p < .01$.) No main effect of Accent was observed, ($F_1(1,20) = 2.64, p = .11$; $F_2(1,68) = .43, p < .51$.) The Truth Value x Accent interaction similarly was nonsignificant, ($F_1(1,20) = 1.94, p = .17$; $F_2(1,70) = .67, p = .41$.)

ANOVA applied to the error data according to the same by-subjects and by-items models described above yielded no significant effects. Overall, error rates were low. Error rates for native English accented stimuli were 2.8% and 4.8% in True and False,
respectively. Error rates for non-native English accented stimuli were 3.9% and 4.3% for True and False respectively.

On average, these participants rated the non-native speaker’s accent as 3.58 (s.d. .21) on a scale of 1-7. No participants correctly identified the native language of the speaker as Dutch, though 3 (of 24; 12.5%) guessed German, a language with similar phonological and prosodic structure.

DISCUSSION

Though not statistically robust, the results of Experiment 1c provide limited support for the overall conclusion that native English listeners process native-accented English faster than non-native English, especially as regards False responses. Taken together with the results of Experiment 1b, a possible interpretation is that overall speaker characteristics, for example, clarity (cf. clear speech, Pichney, Durlach & Braida, 1985; Bradlow & Bent, 2002) could be responsible for the effects observed in experiments 1a and 1b. However, attention to the direction of the means in experiment 1c indicates that, though not significant, the accent effect is nonetheless in the expected direction. This in turn allows it to be plausibly inferred that there was an imbalance in the two speaker’s degree of foreign accent in their respective non-native languages. This conclusion is further supported by the estimates of accent strength elicited by the listeners in the two experiments on their respective post-experiment questionnaires.

Discussion: Experiments 1a-1c

Overall, the three experiments in Chapter 2 indicate that non-native listeners can show an advantage for processing foreign-accented speech over native-accented speech,
when the accent is that of the listener’s L1. Experiment 1a presents this evidence, while Experiments 1b and 1c show that the benefit indicated in 1a is not the result of specific speaker characteristics, but rather, a beneficial effect of the speaker’s accent itself.

The lack of a statistically significant accent effect in Experiment 1c, however, prohibits making a strong claim regarding the hypothesis of Experiment 1a. Two aspects of the results provide a plausible explanation. First, although in Experiment 1c the accent effect is not statistically reliable, the means are in the predicted direction, with longer RTs associated with the NNS for both True and False responses. Moreover, this difference is bigger in responses to False items than True items, consistent with the pattern observed in Experiment 1a and 1b. Thus the overall pattern of results is consistent. Potentially, the Dutch accent to which the native English listeners of Experiment 1c were exposed was strong enough to slow participants’ responses as predicted but yet not strong enough to produce a statistically reliable decrement in performance.

This interpretation is consistent with the accent rating data, which indicate an asymmetry in the degree of foreign accent produced by the two speakers in their non-native language. Recall that both groups of native listeners rated the non-native speaker’s degree of accent on a 7-point scale, with 1 assigned to no accent and 7 indicating a strong accent. The Dutch listeners (Experiment 1b) gave the non-native speaker an average accent rating of 4.5, while the American English listeners (Experiment 1c) gave the non-native speaker an average accent rating of 3.58, a .92 difference which was found to be reliable by independent-samples t-test, (t(46) = 3.14, p < .01), indicating that native
listeners did indeed consider the native English speaker (NW) to have a stronger accent in Dutch than the native Dutch speaker (BN) had in English.

To clarify the difference between the NNL of Experiment 1a and the NL of Experiment 1c (who were responding to the exact same stimuli), the subject and item data from the two experiments were submitted to ANOVA including the factors of Experiment, Truth Value, Accent and List. This analysis indicated no main effect of Experiment, \( F(1,40) = .45, p = .5, F(1,136) = 2.35, p = .13 \), indicating that overall, the NNL and the NL responded equally fast (grand mean for Experiment 1a = 657 ms; for Experiment 1c = 604 ms). A main effect of Truth Value was observed, \( F(1,40) = 103.16, p < .01; F(1,136) = 19.21, p < .01 \), confirming that True responses in both experiments were faster than False responses. The interaction of Truth Value and Experiment was significant in analysis by subjects but not by items, \( F(1,40) = 7.79, p < .01, F(1,136) = 1.18, p = .27 \). The three-way interaction of Truth Value, Experiment and List was also significant, \( F(3,40) = 3.81, p < .05; \) no list factor in \( F(2) \). A main effect of Accent was obtained in analysis by subjects but not by items, \( F(1,40) = 4.37, p < .01, F(1,136) = 1.90, p = .17 \). The interaction of Accent and List was significant, \( F(3,40) = 3.87, p < .05 \). Importantly, the interaction of Accent and Experiment was significant in both analyses, \( F(1,40) = 20.91, p < .01, F(2,136) = 5.70, p < .05 \). Additionally, the three-way interaction of Truth Value, Accent and Experiment was

\[^3\] Note that in analysis by items, the List factor is conflated with the Items factor and is therefore not included. Readers may also note that, while no items were dropped from analysis in Experiment 1a, one item was dropped in Experiment 1c. In order to have equal Item Ns in these combined analyses, the same item (no. 5) was dropped from both data sets. Finally, due to the paucity of effects in error rate analyses in these experiments, no further analysis of error rates were conducted intra-experimentally.
significant in analysis by subjects, \(F1(1,40) = 5.87, p < .01\), but only marginal in analysis by items, \(F2(1,136) = 3.39, p = .06\).

On the strength of the Truth Value main effects and interactions reported above, ANOVAs including Experiment as a factor were performed separately on the True and False data from the two experiments, again by subjects and by items. The overall design was Exp (2) and List (4) between-subjects and Accent (2) within-subjects, for each analysis by subjects. Analysis by items included the factors Experiment (2) and Accent (2) both within-items. For the True responses, no main effects were observed; only an uninterpretable Accent x List interaction was obtained in analysis by subjects, \(F1(3,40) = 4.21, p = .01\). The lack of a significant main effect of Experiment for True responses suggests that native and non-native listener groups were equally able to utilize higher-order contextual information to complete the task, reducing the effects of accented speech (beneficial in Experiment 1a, deleterious in Experiment 1c.) A nonsignificant effect of Accent was predicted here as was obtained, given that the accent associated with a benefit was expected to switch as a function of listener group; thus the advantage enjoyed by NNLs in Experiment 1a was canceled out by the (numerical, see Exp 1c; results) advantage NLs had in Experiment 1c; this is further underscored by a nonsignificant interaction of the Experiment and Accent factors. These means can be seen in Figure 4.
In contrast, intraexperimental analysis of the False responses yielded a different pattern of results. The main effects of Experiment and Accent still failed to obtain, indicating equivalent overall performance between the non-native and native listeners in this task. However, a significant Experiment x Accent interaction was observed, $(F1(1,40) = 23.10, p < .01; F2(1,68) = 6.75, p < .05).$ Planned contrasts revealed that this crossover interaction consisted of a significant advantage for the NNS-accented speech by NNL in Experiment 1a ($t_{23} = 4.08, p < .01$) but no significant advantage for the NS-accented speech when processed by the NL in Experiment 1c ($t_{23} = 1.83, p = .08$).
Responses to False Items: Experiment 1a & 1c

Figure 5. Subject means for False responses in Experiments 1a and 1c. Subjects in Experiment 1a were native Dutch listeners, subjects in Experiment 1c were native English listeners. The stimuli were identical, English sentences. Error bars represent standard errors.

Taken together, the comparisons across Experiments 1a and 1c indicate that participants performed equally well generally, as indicated by the lack of main effects of Experiment. This outcome is surprising, given that the participants in Experiment 1a were processing in a second, less-practiced language while the participants in Experiment 1c were operating their native language. One possible explanation centers on the relative amount of exposure to experimental psycholinguistic testing of the two experimental groups and their motivation for participating. The Dutch participants were recruited through the MPI Nijmegen subject pool, and this database is maintained continuously for as long as individuals can still be contacted. As a result, the subjects of Experiment 1a constituted a group which was generally well practiced in making decisions about
linguistic stimuli under time pressure. Furthermore, these subjects were motivated by earning a modest payment, and as a whole approached the experiment with a serious and professional demeanor. The native English participants recruited at the University of Arizona, in contrast, were mainly 1st year students in entry-level psychology courses with little or no prior experience in behavioral testing in psychology or psycholinguistics specifically. Moreover, these individuals were required to participate in experiments to earn course credit. As a result, these subjects were typically less motivated overall and potentially less inclined to optimize their performance.

Given the participant characteristics outlined above, a possible interpretation of the null effects of the Experiment factor in the analysis of Experiments 1a and 1c is that the deleterious effects of operating in a second language served to slow the Dutch participants RTs across the board. As a result, they performed at the same speed as the English participants despite the latter’s natural advantage for operating in their native language.

Notwithstanding the null effects of Experiment, the specifics of the Truth Value factor and the Accent factor do vary, as shown by higher order interactions. In True responses, the overall lack of any significant effects shows clearly that both groups’ use of context was sufficient to overcome effects of accent. For False responses, when context was less predictive, the significant Accent x Experiment interaction highlights the differential effects of accent- significantly faster responses to mutually accented English for Dutch participants but only numerically faster responses to native accented English for native English participants.
Parallel analyses were undertaken with respect to Experiments 1b and 1c. In this case, the comparison was between groups of native Dutch and native English listeners, each operating in their respective L1s. Thus the aim of these analyses was to show that the two groups performed fairly equivalently with respect to the effects of foreign-accented speech. As in the intra-experimental analyses undertaken with respect to 1a and 1c above, in these analyses items which were dropped in either Experiment’s individual analyses were dropped for both Experiments’ analysis by-items here, in order to avoid violation of statistical assumptions required by ANOVA. This totaled two items, numbers 16 (erroneous in Experiment 1b) and 5 (erroneous in Experiment 1c).

Overall ANOVA including Experiment, Truth Value, Accent and List revealed main effects of Experiment ($F(1,40) = 9.82, p < .01$; $F(1,132) = 46.33, p < .01$), interpreted as an overall RT advantage for the participants in Experiment 1b over experiment 1c regardless of Truth Value or Accent interactions. This observation could be related to the comments above regarding the more practiced nature of the MPI Nijmegen subject pool- as the comparison here is between two groups of listeners operating in their native languages, the special nature of the MPI Nijmegen participants resulted in an overall RT advantage. The finding of a significant Experiment x Accent interaction ($F(1,40) = 10.61, p < .01$; $F(1,132) = 4.15, p < .05$) is consistent with this explanation as well and is explored more specifically below.

This ANOVA additionally yielded a significant main effect of Truth Value in both analyses ($F(1,40) = 51.38, p < .01$; $F(1,132) = 13.42, p < .01$), showing that as expected, True decisions were made more rapidly than False decisions in both
Experiments (note also that no Experiment x Truth Value interaction was obtained) and regardless of the accent associated with the token.

However, main effects of Accent ($F(1,40) = 35.81, p < .01$; $F(1,132) = 10.83, p < .01$) and interactions of the Accent and Truth Value variable ($F(1,40) = 22.33, p < .01$; $F(1,132) = 6.43, p < .01$) were obtained, motivating a more refined assessment of effects between the experiments via separate analyses of the True and False responses.

ANOVA on True responses, crossing Experiment (2) and List (4) within and Accent (2) between subjects indicated a main effect of Experiment, ($F(1,40) = 10.52, p < .01$; $F(1,66) = 20.32, p < .01$) indicating that the native Dutch listeners were faster than the native English listeners for True decisions. A significant higher order interaction, Experiment x List x Accent was also significant, ($F(3,40) = 12.08, p < .01$) but this is largely uninterpretable. Importantly, neither a main effect of Accent nor an Accent by Experiment interaction were observed, suggesting that for both groups, task-specific strategies were strong enough to obviate any deleterious effects of foreign-accented speech in True responses. The means associated with this analysis can be seen in Figure 6.
In False responses, the equivalent analysis reveals again a significant main effect of Experiment ($F_1(1.40) = 8.56, p < .01$ and $F_2(1.66) = 26.14, p < .01$), underscoring the observation that the Dutch listeners in 1b were faster overall than American English listeners in 1c. In contrast to the analysis of True responses, however, this analysis revealed a main effect of Accent, ($F_1(1.40) = 46.78, p < .01$; $F_2(1.66) = 13.05, p < .01$) as well as, in analysis by-subject only, an Accent x Experiment interaction, ($F_1(1.40) = 13.15, p < .01$ but $F_2(1.66) = 3.18, p = .08$). Planned comparisons of this interaction show that the Dutch subjects of Experiment 1b responded significantly more rapidly to native-accented speech than to non-native accented speech, ($t_{23} = 8.51, p < .01$) while no
significant difference between accent conditions was observed for the English participants of Experiment 1c ($t_{23} = 1.83, p = .08$). These means are displayed in Figure 7.

Figure 7. Subject means for False responses in Experiments 1b and 1c. Subjects were native Dutch and English listeners, respectively, each group responding in their native language. Error bars represent standard error values.

Taken together, Experiments 1b and 1c produce strong evidence that, when guessing strategies are not a major component of the response profile (as in False responses), the deleterious effects of processing foreign-accented speech can be seen for native listeners of Dutch. The native English listeners in contrast showed this effect only numerically.
The trio of experiments comprising Chapter Two provides support for the main hypothesis of this dissertation, specifically, that mutually-accented speech can be more rapidly comprehended than canonically-accented speech. However, this conclusion is tempered by the outcome of Experiment 1c, where native English listeners indicated only a modest (and statistically unreliable) deleterious effect of non-native accented speech input.
CHAPTER 3: EXPERIMENTS IN ENGLISH AND SPANISH

The experiments in Chapter Three constitute a replication of the experiments in Chapter Two, contrasting a new language (Spanish). The objective was to investigate whether the effects observed in Dutch would also be obtained in a language with different structural and prosodic features. An additional feature of the experiments in Chapter 3 was a shift in subject factors. The listeners in these experiments differed from the listeners recruited above in several ways. Most importantly, these listeners were living and studying in an L2 environment, and were therefore more adapted to processing native American-English speech than the L2-listeners in Experiment 1a. In addition, these listeners indicated earlier acquisition of L2. These two factors were reasonably expected to have a modulating effect on the advantage for processing L1-accented L2 speech observed in Experiment 1a.

Experiment 2a

In Experiment 2a, native Spanish listeners made true/false decisions to sentences presented in English, the participants’ L2. The primary experimental manipulation, as in other experiments reported in this dissertation, was the accent of the speakers recruited to produce the sentences. In experiment 2a, a native speaker of English and a native speaker of Spanish produced sentences in English, resulting in native accented and non-native accented tokens. The prediction, consistent with results reported above, was that these listeners would process the non-native accented stimuli more rapidly than the native accented stimuli. However, listener factors such as residency in an English-speaking environment and relatively early exposure to English as a second language were
hypothesized to modulate this predicted effect. Specifically, it was predicted that these subjects would show a smaller benefit of mutually-accented speech than was observed in Experiment 1a.

METHOD

Materials

The experimental materials consisted of the same 40 English true/false sentences used in Experiments 1a and 1c. These were divided between 36 experimental pairs and 4 practice pairs. Each sentence consisted of a single clause statement, whose truth or falsity could be easily confirmed by real-world knowledge, e.g. (T) A pelican is a kind of bird; (F) A camel is a kind of bird. The sentences ranged from 3 to 8 words in length, and averaged 5.5 words and 7.49 syllables in length. The materials set is given in Appendix A.

Pretesting

Prior to recording, the materials set was presented to 10 native speakers of Spanish for off-line sentence verification in English, to ensure that non-native listeners were familiar enough with the critical vocabulary to perform the experimental task. These participants were recruited through the University of Arizona Center for English as a Second Language program, and were advanced students, aiming to begin coursework in English in the subsequent semester. All of the experimental materials were correctly responded to by 8 or more participants, indicating that these sentences were generally comprehensible to the targeted L2 listener sample.
Recording

As with Experiments 1a–c above, the Experiments 2a–c required two speakers to generate a materials set across the three experiments: A native speaker of English who produced Spanish with an American English accent, and a native speaker of Spanish who produced English with a Spanish accent. For each language, five native speakers were originally recorded speaking the entire materials set in their native and non-native languages.

Native-Spanish speaking participants were recruited through the University of Arizona’s Center For English as a Second Language. As with the participants recruited for pretesting (above), the speakers were advanced students, preparing to take classes in English in the subsequent semester. Five native speakers were recorded, speaking the entire stimulus set in both Spanish and English. The selection of Speaker YM as the NNS was based upon an estimation of a moderate foreign accent in English by the experimenter, and was additionally driven by the speaker’s sharing the particular Mexican-Spanish dialect, broadly Northern, as the non-native listening population was expected to have. YM was 20 years of age and a native of Hermosillo, Sonora, in Mexico. He had informal exposure to English since the age of 6 and had taken English in school with native Spanish-speaking instructors since the age of 14. YM had lived in the US for approximately 2 years but had been taking English instruction through the Center for English as a Second Language for only the current semester.
Five native English speakers with Spanish proficiency were likewise recorded for the English and English-accented Spanish materials. The native English speaker was judged to have a neutral American accent.

After recording, stimuli were transferred to a computer for editing and experimental presentation. NNS tokens were selected first from among the repetitions, based on overall clarity, with earlier (less practiced) productions preferred over later productions. NS tokens were selected so as to be as closely matched in length with the NNS tokens as allowable, when the production or recording quality was not at issue. Durations of both selections were noted for later processing and analysis, and all tokens were equalized for volume by means of RMS-equalization, using Cool Edit software.

Digital expansion and compression procedures were applied to these tokens between accents in order to equate them for length. For these tokens, only 1 difference of less than 1% in length was observed; this token pair was thus not processed by the PSOLA algorithm. The remainder were subjected to expansion or compression procedures. NS tokens (mean length: 1322 ms) were consistently produced more rapidly than NNS tokens (mean length: 1766 ms), by an average difference of 443 ms. As a result, NS utterances were expanded on average by 16.86% (range: -2.5% to 48%) while NNS utterances were contracted on average to 88.02% (range 75.3% to 97.5%). The final average length was 1547 ms.

Participants

Twenty-four native speakers of Spanish were recruited from the University of Arizona community to participate in this experiment, and were compensated with a small
amount of money for their time. These participants were recruited through the Center for English as a Second Language, the Department of Spanish and Portuguese, the Department of Mexican/American Studies, and the Chicano/Hispano Student Association, as well as by fliers. All were undergraduate students, graduate students, faculty, or staff at the University of Arizona, and used English as well as Spanish in daily communication. Subjects who claimed to be bilingual (in this case, defined as “speaking both Spanish and another language since birth”) were excluded, as were participants who spoke European Spanish. Participants’ experience with English was evaluated after behavioral testing by means of a questionnaire and an offline lexical decision test, identical to those used in Experiments 1a and 1b (given in Appendix D); for details see Table 3.

Table 3: Participant L2 Language Experience and Self-Report, Experiment 2a

<table>
<thead>
<tr>
<th>Age</th>
<th>AOA</th>
<th>YOE</th>
<th>R/exp</th>
<th>W/exp</th>
<th>S/exp</th>
<th>LexDec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>21.6</td>
<td>7.1</td>
<td>14.5</td>
<td>6.0</td>
<td>5.8</td>
<td>6.0</td>
</tr>
<tr>
<td>Minimum</td>
<td>17.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>35.0</td>
<td>15.0</td>
<td>30.0</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>

AOA = age of acquisition; YOE = years of experience; R/exp = reading experience; W/exp = writing experience; S/exp = speaking experience; LexDec = percent correct on offline lexical decision test.
Design

The overall design was identical to that of the experiments in Chapter 2. The factor Accent (Native Speaker, NS; Non-Native Speaker, NNS) was crossed with the factor Truth Value (True, T; False, F), and both factors were varied within-subjects. Four lists were generated for counterbalancing of items in conditions, such that no listener heard the both the true and false versions of an item, nor the same item spoken in both accents. Six participants were tested on each list. Items in each list were pseudorandomized to ensure that no more than three consecutive instances of a particular crossing of stimulus factors were present. Within each list, items were presented in the same pseudorandom order for each subject.

Procedure

Subjects were tested individually in sound-attenuated booths. Presentation of experimental materials and concomitant reaction time measurements were controlled by a computer running DMDX experimental software (Forster & Forster, 2003) Stimuli were presented over headphones at a comfortable listening level, approximately 56 dB SPL. Subjects were given instructions onscreen in English, which stressed the importance of accuracy and speed. Participants responded via a button box with their dominant hand. The reaction time experiment took about 10 minutes. After behavioral testing, participants filled out the language background questionnaire and an offline lexical decision test in order to assess English skills, details of which were reported in Table 3, above. The entire experiment, including debriefing, took about 25 minutes.
RESULTS

Recorded RTs were first adjusted to reflect measurement from the offset of the sentence. Responses were evaluated for accuracy; RTs associated with incorrect responses were dropped from analysis. Reaction times less than –250 ms and greater then 2500 ms were similarly treated as errors. Evaluation of Subject and Item accuracy on these data indicated an overall mean error rate for subjects of 8.68% (maximum 16.67%, 1 subject), no subjects were dropped from analysis. Analysis of Item errors revealed two items (No. 19 and 35; see appendix A) with error rates above 25% (both 29.7%), which were dropped from analysis. The resulting overall error rate for Subjects was 7.4%. Means based on 24 subjects and 34 items can be seen in Figure 8.

These RT data were evaluated by ANOVA with Subjects as the random factor (F1) and items as the random factor (F2). In the analysis by subjects, Accent and Truth Value were within-subjects, and the between-subjects factor of List was included. In the analysis by items, Accent was treated as a within-items factor while Truth Value varied between-items.

This analysis yielded a main effect of Truth, \( F1(1,20) = 31.00, p < .01; F2(1,66) = 14.42, p < .01 \), indicating that ‘True’ responses were faster than ‘False’ responses. Additionally, a main effect of Accent was observed, \( F1(1,20) = 6.16, p < .05; F2(1,66) = 4.44, p < .05 \), indicating faster responses to native accented sentences than to non-native accented sentences. The interaction of Truth Value and Accent was nonsignificant, \( F1(1,20) = 1.42, p = .24; F2(1,66) = .65, p = .42 \). A significant three-way interaction of
the Truth, Accent and List factors was also observed, limited to the analysis by subjects, 

\(F(3,20) = 6.03, p < .01\). No other main effects or interactions were significant.

![Experiment 2a: Spanish Listeners](image)

*Figure 8.* Mean subject reaction times for conditions in Experiment 2a. Subjects were native Spanish listeners, sentences were in English. Error bars represent standard error values.

Error data were also submitted to ANOVA with both subjects (F1) and items (F2) as random factors. In this analysis, only the factor Accent was found to vary significantly, in analysis by-items, \(F(2,66) = 4.30, p < .05\), but only a trend towards significance in analysis by-subjects, \(F(1,20) = 3.40, p = .08\), indicating that evaluation of NS sentences was statistically more accurate than evaluation of NNS sentences, consistent with the RT advantage described above. No other main effects or interactions were observed for error data. The error rates for native English accented stimuli were
3.9% and 7.6% for True and False, respectively. The error rates for Spanish accented stimuli were 8.4% and 9.7% for True and False, respectively.

DISCUSSION

The results of Experiment 2a show clearly that for these Spanish listeners, native accented sentences were both more quickly evaluated as well as more accurately evaluated than non-native accented sentences. This observation is in contrast to the prediction, as well as to the outcome of Exp 1a. A further difference can be seen in the modulation of the Accent effect by Truth Value. In Experiment 2a, to the extent that the Truth Value factor modulated the Accent effect, it did so more for the true responses (133 ms difference between NS and NNS) than for false responses (60 ms difference), though this asymmetry was not statistically significant. In Experiment 1a, the reverse pattern was observed. Possible interpretations of this reversal will be presented in the General Discussion, below.

Experiment 2b

In experiment 2b, native listeners of Spanish made speeded True/False decisions to sentences presented in Spanish. Importantly, listeners heard sentences produced by a native speaker of Spanish and a native speaker of American English, thus constituting native accented and non-native accented materials. The prediction, consistent with experiment 1b above as well as prior research (e.g., Munro & Derwing, 1995), was that the utterances produced by the native speaker would be more accurately and rapidly evaluated than those of the non-native speaker.
METHOD

Materials

The experimental materials consisted of 40 true/false sentence pairs, generated by the experimenter. The sentences were the translation equivalents of the materials used in Experiment 2a (and as well in Experiment 1a and 1c). Each sentence consisted of a single clause statement whose truth or falsity could easily be confirmed by real-world knowledge, e.g. (T) Un pelícano es un tipo de pájaro (A pelican is a type of bird); (F) Uno camello es un tipo de pájaro (A camel is a type of bird.) Sentences ranged from 4 to 10 words in length and averaged 6.3 words in length. The complete set of materials is provided in Appendix B.

Prior to recording, Spanish materials were submitted to native speakers of Spanish for offline sentence verification in Spanish, in order to establish the accuracy of the translations as well as to verify cross-linguistic and cultural validity. Ten native speakers of Spanish recruited through the University of Arizona Center for English as a Second Language program provided these data. All items were responded to correctly by nine or more participants, with the exception of two sentence pairs, which were used as practice items. Two further pairs containing proper nouns were used as practice items. Thus the total stimulus set consisted of 36 pairs of t/f sentences and four pairs of practice t/f sentences.

Recording

Stimulus recording procedures were identical and contemporaneous with the procedures outlined in Experiment 2a. The same speakers were used, with a reversal in
the speaker associated with native accented and non-native accented speech. Specifically, speaker YM (L1: Spanish) produced native accented Spanish while speaker JM (L1: English) produced non-native accented Spanish. Informal judgments by native listeners of Spanish indicated that speaker YM produced Spanish with a northern Mexican accent, while speaker JM produced Spanish with a mild but identifiable American English accent.

As with the previous materials sets, the Spanish materials were equated for length between accents using the PSOLA expansion/compression algorithm in Praat. Prior to this manipulation, NS items averaged 1783 ms in length; NNS items averaged 2058 ms in length, a difference of 274 ms. Seven item pairs were found to have less than 1% difference between their lengths and were thus unchanged. The remaining NS utterances were expanded on average by 7.9% (range –8% to 27%) of their original length and remaining NNS items were compressed on average to 93.8% of their original length (range 82% to 109%). The final average length was 1921 ms.

Participants

Twenty-four native speakers of Spanish were recruited from the University of Arizona community to participate in this experiment. Listeners were compensated monetarily for their participation. Participants were recruited as in Experiment 2a, though relevant departments and student associations. Although these subjects were listening in their native language, English proficiency and background information was collected as in Experiment 2a for the purpose of ensuring parity between subject groups, these data are presented in Table 4. Bilingual subjects, defined by self-report of
contemporaneous acquisition of Spanish and any other language, were replaced.

Similarly, native listeners of European Spanish were replaced.

**Table 4: Participant L2 Language Experience and Self-Report, Experiment 2b**

<table>
<thead>
<tr>
<th>Age</th>
<th>AOA</th>
<th>YOE</th>
<th>Self-report English Experience</th>
<th>LexDec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>R/exp</td>
<td>W/exp</td>
</tr>
<tr>
<td>Mean</td>
<td>27.8</td>
<td>10.8</td>
<td>17.0</td>
<td>5.9</td>
</tr>
<tr>
<td>Min</td>
<td>18.0</td>
<td>4.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Max</td>
<td>53.0</td>
<td>22.0</td>
<td>43.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>

AOA = age of acquisition; YOE = years of experience; R/exp = reading experience; W/exp = writing experience; S/exp = speaking experience; LexDec = percent correct on offline lexical decision test.

**Design**

As in all the experiments reported here, the experimental design of Experiment 2b consisted of two factors—Truth and Accent—fully crossed within subjects and between items, requiring 4-list item-counterbalancing scheme. Pseudorandomization parameters were identical to prior experiments, and each list was presented in the same experimental order to each participant. Six participants heard each list.

**Procedure**

The experimental procedure was identical to those of the previous experiments undertaken in the University of Arizona psycholinguistics lab. Participants were tested individually in sound-attenuated booths containing a monitor, headphones and a labeled response button box. Presentation of experimental materials and reaction time
measurement was again controlled by a computer running DMDX experimental software; stimuli were presented at a comfortable listening level, approximately 56 dB SPL. Subjects were given instructions on-screen in Spanish. The instructions were a close translation of the instructions used for the English language experiments. The reaction time experiment took approximately 10 minutes. After behavioral testing, participants responded to a language background questionnaire querying their acquisition and use of English. Participants also completed an offline English lexical decision test as a means of quantifying their English proficiency. These participants were also asked to rate the degree of foreign accent of the NNS and to guess the speaker’s native language.

RESULTS

Recorded RTs were first adjusted to reflect measurement from the sentence offset. Responses were evaluated for correctness, RTs associated with incorrect responses were dropped from analysis. RTs outside –250 and 2500 ms were similarly treated as errors, together these totaled 8.22% of responses. Evaluation of Subject and Item error rates was conducted on these remaining data. Subject maximum % error was 16.67% (1 subject) and no subjects were dropped from analysis. Two items (19 and 31; See Appendix C) were found to have unacceptably high error rates (both 29%); these were dropped from analysis. Final subject mean error rate based on 34 items was 6.99%.

RT data were submitted to statistical analysis in accordance with the design employed throughout this dissertation. Specifically, Subject (F1) and Item (F2) analyses were conducted on both RT and % Error means. Critical factors were Truth Value, Accent and List analysis by-subjects; Truth Value and Accent for analysis by-items.
These analyses revealed a main effect of Truth Value, \( F(1,20) = 9.57, p < .01; \)
\( F(1,66) = 4.29, p < .05 \), indicating that responses to True items were faster than
responses to False items (see Figure 9, below). The main effect of Accent was significant
in the analysis by-subjects, \( F(1,20) = 4.68, p < .05 \), but indicated only a trend in
analysis by-items, \( F(1,66) = 2.88, p = .09 \), with the means indicating a faster response
to NS than NNS trials. Two significant higher-order interactions were also observed, in
analysis by subjects alone: Accent x List, \( F(3,20) = 4.82, p < .05 \) and the complex
Truth Value x Accent x List interaction, \( F(3,20) = 3.56, p < .05 \).
Analysis of subject’s error data was also conducted, according to the parameters used in the analysis of RT data. These analyses yielded a similar pattern of results in analysis by subjects ($F_1$), but no significant results in analysis by items ($F_2$). Both analysis revealed a trend towards significance for the main effect of Truth Value: ($F_1(1,20) = 4.02, p = .06$ and $F_2(1,66) = 2.87, p = .09$). A main effect of Accent, ($F_1(1,20) = 6.12, p < .05$ but $F_2(1,66) = 2.88, p = .09$) and an Accent x List interaction, ($F_1(3,20) = 5.49, p < .01$ and no $F_2$ due to lack of List factor in Items analysis) were also significant. The mean error rates for the native Spanish stimuli were 4.6% and 6.2% in True and False respectively. The mean error rates for the non-native Spanish accented stimuli were 6.0% and 11.4% for True and False, respectively.

Accent rating data indicated an average rating of the NNS’s accent of 4.04 (s.d. .21) on a 7-point Likert scale, with 7 indicating a heavy accent. Nearly all participants correctly identified the native language of the NNS as American English, N = 22 (92%).

**DISCUSSION**

The results of Experiment 2b, though not robust, indicate that native Spanish listeners were faster to respond to NS than NNS Spanish items. Clearly, however, for these listeners, the NNS trials were not as difficult to process as the NNS trials were in Experiment 1b. It is possible that the native-English speaker chosen for Experiments 2a–2c had a more native-like accent in Spanish than the non-native speaker of English in Experiments 1a–1c had in Dutch. It is also possible, however, that the listeners in Experiment 2b had more experience listening to English-accented Spanish than the Dutch
subjects in Experiment 1b had listening to English-accented Dutch. These two possible explanations will be discussed in more detail below.

Experiment 2c

In Experiment 2c, native listeners of English made speeded true/false decisions to sentences presented in English. The accent of the speaker was the primary manipulation; listeners were exposed to native-accented English and Spanish-accented English sentences, i.e. native accented and non-native accented materials. The predication, based on Experiment 1c and other studies (Munro & Derwing, 1995), was that listeners would be faster and more accurate in processing native accented speech than in processing non-native accented speech.

METHOD

Materials

The experimental materials consisted of the same set of 40 English true/false sentences employed in Experiment 2a. The same files were used.

Recording

Recording procedures were identical and coterminous with those of Experiment 2a.

Participants

Twenty-four Native speakers of English were recruited from the University of Arizona Psychology Department subject pool, and were given course credit for their participation. None had extensive experience with the Spanish language.
Design

The experimental design was identical to that of Experiment 2a.

Procedure

Experimental procedures were similar to those of the previous experiments. These participants did not respond to the language background questionnaire or the lexical decision test, but were asked to provide an accent rating for the NNS on a 7-point Likert scale (7 indicating a heavy accent), and to guess the origin of the accent.

RESULTS

Recorded RTs were first adjusted to reflect measurement from the sentence offset. Responses were evaluated for correctness and RTs associated with incorrect responses were dropped from analysis. Scores outside of the −250 to 2500 ms range were similarly excluded. Together these accounted for 6.83% of the data. Next, individual subjects and items were evaluated for accuracy. The maximum subject error was 13.89%, and no subjects were excluded from analysis. One item (No. 26) had an error rate exceeding 25.0% (29.0%) was therefore excluded from analysis. Final mean error rate for Subject N = 24 and Item N = 35 was 6.19%

RT data were submitted to an ANOVA with Subjects as the random factor (F1), and with Items as the random factor (F2), as in prior experiments. These analyses revealed a main effect of Truth Value in analyses by subjects, \( F(1,20) = 7.43, p = .01 \), which was only marginal in analysis by items, \( F(1,68) = 3.50, p = .06 \). A main effect of Accent was found to be robust in both analyses, \( F(1,20) = 64.53, p < .01; F(1,68) = 21.45, p < .01 \). Analysis by-subjects also revealed significant interactions of Accent x
List and Truth Value x Accent x List, \((F1(1,20) = 3.82, p < .05\) and \(F1(1,20) = 7.51, p < .01\), respectively). The condition means are reported in Figure 10.

Experiment 2c: English Listeners

![Figure 10](image)

*Figure 10.* Mean subject reaction times for conditions in Experiment 2c. Subjects were native English listeners, sentences were in English. Error bars represent standard error values.

ANOVA applied to by-subject and by-item error rates indicated a main effect of Accent, \((F1(1,20) = 13.81, p < .01; F2(1,68) = 4.07, p < .05)\), indicating that responses to the NS were more accurate (4.04%) than responses to the NNS (9.51%). The error rates for native accented items were 5.3% and 2.8% for True and False conditions, respectively. The error rates for the non-native accented items were 9.3% and 9.7% for True and False conditions, respectively.
The mean accent rating associated with the NNS in this experiment was 5.25 (s.d. .17), on the 7-point scale, and participants were generally accurate in identifying the native language of the speaker as Spanish, N = 21 (88%).

DISCUSSION

The results of this experiment closely follow predictions and serve further to aid interpretation of the results of Experiment 2a. Importantly, native English listeners responded similarly to the native Spanish listeners tested in Experiment 2a when evaluating native accented and non-native accented English. They were statistically faster, as well as more accurate, when evaluating native-accented than nonnative-accented speech.

Discussion: Experiments 2a–2c

Taken together, the results of experiments 2a–2c indicate that for native Spanish listeners with broad exposure to English, sentence verification times are faster when produced in a native accent than in a non-native accent. That is, no benefit of mutually-accented speech is observed for the NNL subjects in Experiment 2a. The follow-on experiments indicated that similar listeners do show the expected advantage for native-accented speech when processing in L1 (Spanish; Experiment 2b) and that native listeners of English processing the same stimuli used in Experiment 2a showed a similar advantage for processing native-accented speech over foreign-accented speech (Experiment 2c).
First, consider the outcomes of Experiment 2a and 2c together. This contrast consists of NNL and NL's performance on the identical stimuli, Native- and Spanish-accented English sentences.

These data were first subjected to an ANOVA including the variables of Experiment (2), Truth Value (2), List (4) and Accent (2), in order to examine the effects of listener group (the Experiment variable) across the two experiments. Note that in order to equate the number of items between the two experiments, all the items dropped in either experiment were dropped in this analysis. Specifically, items 19 and 35 were dropped (erroneous in Experiment 2a) and item 26 (erroneous in Experiment 2c). This analysis revealed a main effect of Experiment, \( F_1(1,40) = 10.16, p < .01; F_2(1,128) = 31.85, p < .01 \), indicating that the NL in Experiment 2c were significantly faster overall than the NNL in Experiment 2a. A main effect of Truth Value was also obtained, \( F_1(1,40) = 37.22, p < .01; F_2(1,128) = 16.26, p < .01 \) indicating that overall, listeners were faster to make True responses than False responses. However, the strength of this effect for the two listener groups was not equal, as indicated by the significant interaction of the Truth Value x Experiment factors, \( F_1(1,40) = 8.22, p < .01 \) and \( F_2(1,128) = 4.02, p < .05 \). Main effects of the Accent variable were also obtained in both analysis by-subjects and by-items, \( F_1(1,40) = 38.96, p < .01 \) and \( F_2(1,128) = 18.97, p < .01 \).

The Experiment x Truth Value interaction, as well as main effects of the Accent factor, were evaluated in separate ANOVAs for each Truth Value condition. For True responses, an ANOVA crossing the Experiment (2), List (4) and Accent (2) factors was performed. This analysis showed no significant effect in analysis by-subjects and a
significant effect in analysis by-items for the Experiment factor, \((F1(1,40) = 3.80, p = .06; F2(1,64) = 8.08, p < .01)\), indicating that within True responses, the NNL in Experiment 2a were somewhat slower overall than the NL in Experiment 2c.

Additionally, a main effect of Accent was obtained, \((F1(1,40) = 44.70, p < .01; F2(1,64) = 12.25, p < .01)\), showing that for both listener groups, native accented speech was easier to evaluate than non-native accented speech. The Accent x Experiment interaction was not significant in either analysis, \((F1(1,40) = 1.51, p = .22; F2(1,64) = 1.56, p = .21)\). The lack of a significant Experiment x Accent interaction in either analysis suggests that the deleterious effect of the non-native accented speech was similar for both listener groups, when the role of higher-order information was encouraged. The means entered into this ANOVA are displayed below in Figure 11.
Within False responses, an equivalently designed ANOVA revealed a main effect of Experiment, \((F1(1,40) = 15.18, p < .01; F2(1,64) = 24.77, p < .01)\), indicating that again, NL were faster than NNL. Similarly, a main effect of Accent, \((F1(1,40) = 13.21, p < .01; F2(1,64) = 6.77, p < .01)\), showed that responses to NS were faster than responses to NNS overall. This analysis revealed no significant interaction of the Experiment and Accent factors, \((F1(1,40) = 3.27, p = .08; F2(1,64) = 1.22, p = .27)\). This outcome suggests that even when strategic factors had less of an impact on the decision process, the NNL were no more disadvantaged by NNS than the NL were. Figure 12 shows the means of the conditions in this analysis.
Responses to False items: Experiments 2a & 2c

In summary, the detailed comparisons of performance between Experiment 2a and 2c suggest that the two subject groups performed broadly equivalently with respect to accent, while the NNL in Experiment 2a were slower overall than the NL in Experiment 2c. The NNL were additionally somewhat less able to successfully incorporate contextual information in responding, and were less impeded by the NNS accent in False than True responses.

An equivalent series of analyses were conducted on the data from Experiments 2b and 2c, in order to compare the responses of two different native speaker groups on an intralingually consistent materials set. As in prior analyses, item sets dropped in either original experiment were dropped from both experimental data sets, three items total.
comprising two items from Experiment 2b (items 19 and 31) and one item from Experiment 2c (item 26). An ANOVA including the factors of Experiment (2), Truth Value (2), List (4) and Accent (2) yielded a significant main effect of the Experiment factor, \(F(1,40) = 5.53, p < .05; F(1,128) = 17.69, p < .01\), indicating that the NL of English (Experiment 2c) were faster overall than the NL of Spanish (Experiment 2b). In addition, a main effect of Truth Value was also obtained, \(F(1,40) = 16.96, p < .01, F(1,128) = 7.72, p < .01\) showing that responses to True sentences were faster than responses to False statements. The Truth Value x Experiment interaction was not significant by-subjects or by-items, suggesting that the size of the Truth Value effect was similar for both listener groups. Substantial Accent effects were observed, \(F(1,40) = 50.58, p < .01; F(1,128) = 17.69, p < .01\) as well as Accent x Experiment interactions, but no significant Accent x Truth Value effects were discovered. Separate ANOVAs on the True and False responses were subsequently undertaken.

ANOVA on True responses revealed a significant main effect of Experiment, \(F(1,40) = 5.41, p < .05; F(1,64) = 9.71, p < .01\), showing as in the overall ANOVA above, an RT advantage for the NL in 2c (Native English) over the NL in 2b (Native Spanish). This analysis additionally indicated significant main effects of Accent \(F(1,40) = 26.22, p < .01\) and \(F(1,64) = 8.00, p < .01\), showing that more rapid decisions were made to the canonically-accented NS than the foreign-accented NNS. The Accent effects were not as strong for the Spanish participants in Experiment 2b as for the English participants in Experiment 2c, as was indicated by a reliable Accent x Experiment interaction, \(F(1,40) = 16.23, p < .01; F(2,64) = 6.15, p < .05\) and planned
contrasts for each level indicating no effect for the Spanish participants \( t(23) = .53 \) but a robust effect for participants in Experiment 1 \( t(23) = 5.14, p < .01 \). The means corresponding to this analysis are displayed in Figure 13, below.

![Figure 13](image.png)

Figure 13. Subject means for True responses in Experiments 2b and 2c. Subjects were native Spanish and English listeners, respectively, each group responding in their native language. Error bars represent standard error values.

An analysis of the False responses led to a main effect of the Experiment factor \( (F1(1,40) = 4.44, p < .05; F2(1,64) = 8.55, p < .01) \), confirming that RTs across levels of Accent were faster in Experiment 2c than in 2b. A main effect of Accent \( (F1(1,40) = 20.44, p < .01; F2(1,64) = 9.69, p < .01) \) indicated an RT advantage for NS sentences over NNS productions in both Experiments. In contrast to responses to True statements, no Accent x Experiment interaction was observed for False responses, suggesting that the
deleterious effects of foreign-accent were similar in both experiments. Note however that the difference between the two conditions is clearly numerically larger in Experiment 1c ($\Delta = 178$ ms) than in Experiment 1b ($\Delta = 60$ ms). These means are presented in Figure 14.

![Responses to False items: Experiments 2b & 2c](image)

**Figure 14.** Subject means for False responses in Experiments 2b and 2c. Subjects were native Spanish and English listeners, respectively, each group responding in their native language. Error bars represent standard error values.

In summary, as was predicted by hypothesis and prior research, a similar pattern of results was found for both NL experiments. Native listeners responded faster to True than False statements, indicating the expected use of context to facilitate decision-making. NLs also displayed and faster responses to canonically accented speech than to foreign accented speech. Interestingly, the lack of a significant Accent x Experiment
interaction for the False but not the True responses suggests that when efficiency in bottom up processing was more critical to making a rapid responses, the deleterious effects of accent were similar for both groups of listeners. When use of context was encouraged, however, Spanish listeners showed less strong negative effects of foreign accented speech than the English listeners, suggesting more efficient comprehension of the incoming signal leading to a stronger contribution of context to the response. However, English listener’s difficulty with the bottom-up processing of the input potentially delayed and reduced the contribution of ongoing contextual processing and assimilation to their responses. Further interpretation of the results of the Spanish/English experiments 2a-2c with respect to the outcomes of Experiments 1a-1c is contained in Chapter 4.
CHAPTER 4: GENERAL DISCUSSION

Summary Of Findings

The major finding of the research in this dissertation is that non-native listeners can benefit from processing mutually-accented non-native speech, however, speaker, listener and environmental factors play a modulating role in this effect. Secondary findings include the observation that non-native listeners can effectively employ contextual strategic guessing when performing the SVT in their L2, and the replication and extension of previous findings for native listeners indicating slower verification times for non-native accented than native accented stimuli when performing the SVT.

The results of Experiment 1a support the primary prediction. In this experiment, native Dutch listeners correctly verified English sentences faster when they were produced with a light Dutch accent than in a canonical American English accent, in contrast to native English listeners who showed no difference for the same tokens in Experiment 1c. Non-native participants' accuracy in Experiment 1a was numerically, though not statistically, complimentary to the reaction time data, showing lower error rates for non-native accented than native accented speech.

The finding of a main effect of Truth Value for the Dutch participants of Experiment 1a provides novel evidence that non-native listeners are able to effectively draw upon higher-order guessing strategies when operating in their second language. Furthermore, the trend towards significance in the interaction of the Truth Value and Accent for these subjects suggests that the beneficial aspects of mutually-accented speech
contributed more strongly to response latencies when the role of such cognitive guessing strategies was limited, as in the case of False responses.

However, the non-native listeners in Experiment 2a (L1: Spanish) indicated a different pattern of results. These participants showed faster reaction times to correctly verify canonically-accented English than moderately Spanish-accented English, similar to native English participants responses to the same tokens in Experiment 2c. Crucial differences in the amount of exposure to English between the non-native listeners in Experiments 1a and 2a plausibly account for the differing pattern of results. Specifically, the non-native listeners in Experiment 2a had significantly younger mean age of acquisition for L2, significantly longer amounts of exposure to L2, and significantly higher self reports of proficiency in English than the NNLs in Experiment 1a. Moreover, the NNLs in Experiment 2a were more strongly in a bilingual (or even monolingual English) mode than the participants in Experiment 1a. The participants in 2a were tested as they lived and studied in an L2 environment (The United States), while the participants in Experiment 1a were tested in their native-language environment (The Netherlands).

The results of Experiment 1b, 1c, 2b and 2c effectively extend previous findings which show that non-native accented speech is less comprehensible than native accented speech (Munro & Derwing, 1996). Experiments 1b and 2b show this for native listeners of Dutch and Spanish, respectively, extending previous findings to previously untested language and accent contrasts. Experiments 1c and 2c underscore the importance of speaker characteristics in the strength of this effect, for native English listeners.
In addition, the relatively limited experience with American English of the subjects in 1a coupled with environmental factors (testing in Holland) may have led to the pattern of results. Experiment 2a showed a stable pattern but also indicated an advantage for canonically-produced L2 speech over L1-accented L2 speech for the Spanish participants. Here, the control experiments also indicate an asymmetry in the degree of accent displayed by the two speakers in their non-native language, though less so than in Experiments 1a and 1c. Here again, as in Experiment 1a and 1b, the site of testing (in this case, the United States) as well as experience with English may have played a role.

The participants relative experience using the second language tested in these experiments (English for native Dutch and Spanish listeners, respectively) is likely to have played a role in the different outcomes of Experiment 1a and 2a. Because the subject groups were shown to differ on factors related to English use, it seems likely that for the relatively inexperienced listeners, English produced in a matching accent was easier to process. The Spanish participants, in contrast, were more experienced at processing English and were able to perform much like native English listeners.

**Signal Detection Analyses**

In order to more thoroughly characterize the difference in L2 experience between the L1 listener groups in Experiments 1 and 2, a signal detection theory analysis was carried out on the offline lexical decision data. These analyses confirmed that the native Spanish listeners were more familiar with the English vocabulary contained in the lexical
decision tests than the Dutch subjects, but that both groups were high in sensitivity and neither were biased in responding.

The native Dutch participants in Experiment 1a as a group revealed an overall $d'$ score of 1.68, calculated from all participants’ responses. Individual $d'$ scores were calculated using a corrected score for participants who had no misses or false alarms (N=8) by substituting .5 errors for 0 errors, allowing a $d'$ calculation to be made. Using the corrected scores, the overall group $d'$ was 1.63, with a range from .42 to 3.92. The mean individual $d'$ score was 1.84.

The native Dutch participants in Experiment 1b as a group revealed a comparable overall $d'$ score of 1.45, calculated as above from the responses of the entire group. Applying the same correction as above for participants with no misses or false alarms (N = 10) yielded an overall group $d'$ of 1.37, ranging from .34 to 2.56 and averaging 1.50.

The same procedures were next applied to the lexical decision data from the native Spanish participants. The data from Experiment 2a indicated an overall group $d'$ score of 2.20. Applying the correction procedures outlined above for participants with no misses or false alarms (N = 9) led to an overall group $d'$ score of 2.16, ranging from 1.01 to 3.89 and averaging 2.48.

Native Spanish participants’ lexical decision data from Experiment 2b were likewise analyzed. Uncorrected group data indicated an overall $d'$ of 1.98; correction procedures previously detailed yielded a $d'$ score of 1.95 (correction applied to N = 6 participants). The corrected scores ranged from .47 to 3.92 with a mean of 2.36.
The $d'$ data were further analyzed by univariate ANOVA, contrasting the participant’s native language and the language of the main experiment between subjects. This analysis revealed a main effect of native language, $(F(1,95) = 21.69, p < .01)$, indicating that the native Dutch participants had significantly less sensitivity than the native Spanish participants (mean $d'$ scores: 1.66 and 2.41 respectively). The experiment language factor showed no statistically significant effect, $(F(1,95) = 2.09, p = .15)$, nor did the interaction between native language and experiment language, $(F(1,95) = .42, p = .51)$. Thus group differences in sensitivity as a function of experimental language played no role in overall sensitivity—having just responded in the SVT did not lead to better performance on the lexical decision test.

These results support the interpretation that the Spanish subjects overall were more proficient in English than the Dutch subjects. Recall that Spanish subjects made significantly fewer errors on the lexical decision test than Dutch subjects. The SDT analysis underscores the conclusion that differences in performance between the two groups were caused by differences in sensitivity, and were not purely a function of biased strategic responding.

This interpretation is consistent with other studies indicating an impact of age of L2 acquisition on non-native speech perception (e.g., Florentine, 1985; Mayo, Florentine and Buus, 1997). For example, Mayo et al. found that Spanish-speaking listeners who learned fluent English before age six differed from participants who had acquired L2

---

4 As participants completed the RT experiment prior to the offline lexical decision test, the language of the experiment itself might reasonably be expected to affect performance on the lexical decision test, justifying its inclusion as a factor.
after puberty with respect to the negative impact of noise on comprehensibility. Their results moreover indicated that age of acquisition more than amount of experience was responsible for the observed difference in performance. In addition, Mayo et al. (1997) showed that age of acquisition similarly impacted participants' ability to integrate across-word contextual information, as indexed by reduced effects of predictability in their sentence-final target words.

**Multiple Regression Analyses**

As a means of exploring the relationship between subject language background variables and performance on the SVT, regression analyses were performed on the data from Experiments 1a, 1b, 2a and 2b. Recall that non-native participants in Experiment 1a were observed to respond to mutually-accented speech in L2 more rapidly than to unaccented speech, while the non-native participants in Experiment 2a showed the opposite pattern: faster responses to native accented speech than to mutually-accented speech. In contrast, listeners performing the SVT in their native language in Experiments 1b and 2b indicated reliably more rapid decisions to native accented than to non-native accented speech. The correlation analyses sought to determine whether this pattern of results was related to the participants’ experience with English, as defined by responses on the language background questionnaire and performance on the lexical decision test. English was either the language of the experiment (in Experiment 1a and 2a) or the origin of the accent when the language of testing was Dutch (Experiment 1b) or Spanish (experiment 2b), and so participants’ relative familiarity with English could be expected to affect response times in any of the four experiments.
Non-native Listener Analyses

First, subject variables from experiment 1a were entered into a linear regression with the amount of non-native speaker accent advantage for responses to false items as the dependent variable. All the subject factors (Age of Acquisition, Years of Experience, Self report of Reading, Writing and Speaking Experience, and Error Score on the Lexical Decision test) were entered into the model as there were no a priori theoretical reasons to restrict or order them. This model yielded an overall $R^2$ coefficient of .20, indicating little variance accounted for, and was moreover not significant, $p = .66$. Inspection of the $p$-values associated with each of the factors entered into the model revealed that none accounted significantly for any amount of variance in the accent advantage. This outcome provides no evidence for a relationship between subject’s experience with English and the size of the observed advantage for mutually-accented speech input.

Parallel regression analyses were performed using the accent advantage observed for true responses. The same regression model was employed as for the false responses above. This model indicated no overall effect: $R^2 = .28$, $p = .39$, and no individual factors indicated a significant effect. Again the lack of a positive and significant correlation here provides no evidence for the possibility that background exposure to English played a role in these subjects’ response times to English- vs. Dutch-accented English speech input.

For the native Spanish-listening participants in Experiment 2a, the same regression analyses were performed. Regression of the subject factors along the accent advantage for false responses in this case yielded no significant effects, $R^2 = .38$, $p = .17$. 
The same outcome was observed in regression analysis of the subject factors using native speech accent advantage for true responses. Here the observed $R^2$ value was .33, $p = .27$.

Overall, for the non-native listener groups taken together, there appears to be little evidence of a modulation of accent effects of background experience as measured by the instruments in this study.

Native Listener Analyses

An identical sequence of multiple regression analyses was conducted to examine the potential relationship between experience and the strength of observed detrimental effects of accented speech for the NL groups in Experiments 1b and 2b. As the participants were native listeners (in contrast to the regression analyses described above), these analyses were aimed at determining whether more experience with English would reduce the penalty associated with English-accented speech, leading to significant correlations between the language experience variables and reaction times.

Regression analyses of the false responses for native listeners in Experiment 1b were calculated with the model outlined for non-native listeners above. An overall $R^2$ value of .23 was observed but was not significant, $p = .54$.

True responses lead to a digression from the broader pattern of results, showing a significant overall model $R^2$ of .50 at $p < .05$. The Age Of Acquisition factor was found to be associated with this outcome, leading to a meaningful interpretation: As age of acquisition of English for these non-native Dutch listeners increased, the difference in processing speed between the native- and English-accented Dutch SVT responses decreased. That this outcome was observed for responses to True but not False items
suggests in particular that participants' experience processing English led to more successful comprehension and implementation of predictive strategies as regarded the sentence-final words.

The multiple regression analyses applied to native Spanish listener's data revealed no effects for either the true or false decisions. $R^2$ of .45 was obtained for true decisions, $p = .08$, and analyses of true decisions yielded an $R^2$ of .18 and $p = .72$.

Overall, the outcome of the multiple regressions analyses provides little evidence for a relationship between observed accent advantages (Experiment 1a) or disadvantages (Experiments 1b, 2a & 2b). However, this outcome does not exclude the possibility that a relationship existed but was not readily detectable with the measures employed in the current research. There is wide use of such measures (e.g. questionnaire data on age of acquisition, self-report of proficiency, L2 vocabulary estimation via offline lexical decision) in the literature on second language processing, however, such measures are relatively broad-gauged indices. In the current experiments, these measures sufficiently differentiated the target non-native listener populations with respect to overall experience with English. But, these measures did not systematically pattern with performance on the SVT. It may be the case that finer-grained measures of second-language competency, coupled with more focused measures of comprehension, are needed to sufficiently describe the relations between native and non-native processing addressed in these experiments.
Productive Conversion vs. Form of Stored Representations

One question concerning the locus of the observed effects in these experiments is whether the benefit observed for the less experienced subjects in Experiment 1a is the result of relative efficiency in productively converting the input at the level of phonetic categorization or rather a more precise fit between the input and the stored representation. For example, take the case of identifying the final word in sentence 24(F): *A violin is a tool*. A native speaker of Dutch would produce a non-aspirated initial stop in this location, while English speakers would aspirate the segment. A native Dutch listener might not experience difficulty in categorizing the initial consonant as a /t/, because it categorizes well with the Dutch, though not the English production specifications. In contrast, a native English listener would presumably perceive the initial /t/ as a relatively poor token of /t/ (or even as the phoneme /d/), and this delay in categorization would delay identification at the lexical level, leading to slower RTs in the overall decision.

Alternately, a benefit for native-accented speech could have as its locus the form of the lexical representation(s) initially activated by the input. On this hypothesis, the lexical representations of the adult L2 learner, having been categorized under the influence of the learner’s L1 processing strategies, might be themselves said to be 'accented', that is, representations whose best fit would be typically-accented L2 input.

Preactivation of lexical forms is only a potential feature of the True condition and would not be expected to play a role in responses in the False condition, so it is the True condition where one would expect to see lexical effects emerge most clearly. However, the accent effects in the True condition were generally less robust than in the False
condition. This was argued in Chapter 2 to indicate a kind of floor effect (faster overall response times for True as compared with False responses). In light of this pattern, the possible impact of accented lexical representations is difficult to evaluate from these data.

**FUTURE DIRECTIONS**

There are several factors which merit further investigation given the findings reported above. Chief among these is the question of the extent to which listener factors play a role in modulating the effects of processing mutually-accented speech in L2. The differences between the Dutch and Spanish experiments reported in this dissertation are most plausibly understandable in these terms. However, the attempt to establish a direct relation between L2 experience and SVT performance using the multiple regression tests described above provided only limited success. Thus, the relation between comprehension success as measured with the SVT and language experience is likely a complex one, and evaluation of it will require a breakdown and refinement of the variables used to represent linguistic experience.

Consider, for example, the potential importance of age-of-acquisition of English. As can be seen in Tables 1 and 3, the native Spanish listeners recruited in Experiment 2a had on average an earlier initial exposure to English than the native-Dutch listening participants recruited in Experiment 1a. Age of Acquisition has been shown to be a relevant factor with respect to production and perception of a second language, as indicated by the work of Flege (e.g. Flege, 1995) for one review). In these studies, earlier age of acquisition correlates positively with degree of perceived foreign accent by
native listeners, showing that the earlier one begins to learn an L2, the less predominant
the accent in production.

The results of the current experiment extend this finding to the realm of sentence
comprehension. The participant group with earlier AOA, the Spanish listeners,
performed much like the native listeners in all the native-listener experiments (1b, 1c, 2b
and 2c) that is, they made more rapid sentence verification decisions to native-accented
speech than to non-native accented speech. In contrast, the Dutch listeners, who on
average began learning English at a considerably later age than the Spanish listeners,
seemed to benefit from hearing English as accented by their native language.

While this finding is certainly of interest and importance itself, it leaves open the
question of whether and when this shift from a native to a non-native listening strategy
ever occurs for native Dutch listening participants. Future research could easily address
this issue by targeting native Dutch listening participants with similar early exposure to
English. Based on the evidence presented in these experiments, it could reasonably be
expected that a participant sample such as described might reveal a pattern of responses
more similar to those observed for the more experienced Spanish listeners in Experiment
2a.

However, a second, highly relevant factor could also be usefully explored. In
contrast to considering participant’s absolute degree of knowledge and experience with a
particular L2, consider the potential role of the relative prominence and availability of
that knowledge at the time of testing. As noted, another fundamental difference between
the two subject groups concerns the participants’ living and working environment.
Importantly, while the Dutch participants were tested in a native-language environment (The Netherlands), the Spanish participants were tested in a non-native language environment (The United States). This difference leads to a possible role of the language mode of the participants. Grosjean (2001) has argued that bilingual’s and advanced L2 listeners’ attentional focus on one language or another at any given moment varies along a continuum, ranging from fully monolingual in one language or the other to a mixed or bilingual mode. Environmental context can influence where along this continuum a listener is at any given time. For example, if a listener is involved in a conversation with a speaker who is competent in the same two languages as the listener, code-switching between the two languages is not uncommon, and the interlocutors are likely to be in a mixed, or bilingual mode rather than a purely L1 or L2 mode. On the other hand, if the same listener is in a conversation with someone who is known to only speak one of the listeners’ two languages, the listener is likely to have that language activated more strongly than the other.

With respect to the participants in the current experiments, the Spanish listeners in Experiment 2a were almost certainly in a relatively mixed mode, and quite possibly even giving English somewhat more attention than Spanish. They were living in an English speaking environment, and although they were knowingly recruited for being native speakers of Spanish, they were aware that it was their English skills which were being called upon. In contrast, the Dutch subjects most likely had not been actively using English as extensively prior to testing, and so were predominantly in Dutch mode until engaging in English experimentation.
Therefore, future explorations of the effects observed in this experiment would benefit from one of two possible alternations. The first would be to find a native-Dutch speaking participant group living in the US and with a similar degree of daily use of English as the Spanish participants. Conversely, finding and testing a group of native-Spanish participants in Mexico might yield a pattern of results more similar to those observed for the Dutch subjects in Experiment 1a, namely, an advantage for Spanish-accented English over native-accented English.

In summary, the experiments presented in this dissertation provide evidence that non-native listeners with relatively little L2 listening experience benefit from processing mutually accented speech. In contrast, L2 listeners with extensive L2 experience perform more like native listeners, showing less efficient processing of mutually-accented speech. The control experiments suggest that qualities of the speaker including degree of accent and overall speaking clarity may also interact with comprehensibility and modulate observed beneficial or deleterious effects of accented speech. A further finding is that across six experiments, subjects display effective integration of higher-order guessing strategies leading to a consistent advantage for True over False decisions. Interestingly, this observation holds whether participants are performing the SVT in their native or second language.
APPENDIX A: ENGLISH MATERIALS

Appendix A contains the materials set, in English. Sentence pairs are given in the order False followed by True:

1) A camel is a type of bird/A pelican is a type of bird
2) A tiger has wings/An eagle has wings
3) A cabbage is a type of fruit/A strawberry is a type of fruit
4) A bean is sweet/An apple is sweet
5) A sparrow is a mammal/A cow is a mammal
6) A shirt is a body part/A foot is a body part
7) A shoe has fingers/A hand has fingers
8) A monkey is a type of fish/A tuna is a type of fish
9) A drill is a musical instrument/A piano is a musical instrument
10) A violin is a tool/A hammer is a tool
11) People wear pants on their hands/People wear socks on their feet
12) Some people keep giraffes as pets/Some people keep dogs as pets
13) Most cars run on apple juice/Most cars run on gasoline
14) Denmark is a country in Asia/Spain is a country in Europe
15) A horse has three legs/A horse has four legs
16) Smoking is good for your health/Exercise is good for your health
17) An hour is forty-five minutes/A minute is sixty seconds
18) Milk contains alcohol/Wine contains alcohol
19) People have walked on the sun/People have walked on the moon
20) Some people drink tea with salt/Some people drink coffee with sugar

21) Sometimes elephants eat people/Sometimes tigers eat people

22) A plant needs milk to live/A plant needs water to live

23) A green light means stop/A red light means stop

24) Paper is made from weeds/Paper is made from trees

25) A bicycle is a weapon of war/A tank is a weapon of war

26) Buddhism is a political theory/Hinduism is a religion

27) Spaghetti is a French dish/Spaghetti is an Italian dish

28) A vet can help you when you're sick/A doctor can help you when you're sick

29) Biking is usually slower than walking/Biking is usually faster than walking

30) Children are never afraid of the dark/Children are often afraid of the dark

31) Paper is a good material for buildings/Stone is a good material for buildings

32) Accounting is an art form/Accounting is an occupation

33) June is a winter month/July is a summer month

34) A lamp is a vehicle/A car is a vehicle

35) A chair has a motor/A truck has a motor

36) A boat is a piece of furniture/A table is a piece of furniture

Practice Items:

P1) An orange is crunchy/A carrot is crunchy

P2) A brick is a gem/An emerald is a gem

P3) The capital of England is Manchester/The capital of the US is Washington

P4) Tom Hanks is a famous musician/Tom Hanks is a famous actor
APPENDIX B: DUTCH MATERIALS

Appendix B contains the entire Dutch materials set. Sentence pairs are given in the order False followed by True.

1) Een kameel is een soort vogel/Een pelikaan is een soort vogel
2) Een tijger heeft vleugels/Een adelaar heeft vleugels
3) Een kool is een soort vrucht/Een aardbei is een soort vrucht
4) Een boon is zoet/Een appel is zoet
5) Een mus is een zoogdier/Een koe is een zoogdier
6) Een overhemd is een lichaamsdeel/Een voet is een lichaamsdeel
7) Een schoen heeft vingers/Een hand heeft vingers
8) Een aap is een soort vis/Een tonijn is een soort vis
9) Een boor is een muziekinstrument/Een piano is een muziekinstrument
10) Een viool is een werktuig/Een hamer is een werktuig
11) Mensen dragen een broek aan hun handen/Mensen dragen sokken aan hun voeten
12) Sommige mensen hebben giraffes als huisdier/Sommige mensen hebben honden als huisdier
13) De meeste auto's rijden op appelsap/De meeste auto's rijden op benzine
14) Denemarken is een land in Azië/Spanje is een land in Europa
15) Een paard heeft drie benen/Een paard heeft vier benen
16) Roken is goed voor je gezondheid/Beweging is goed voor je gezondheid
17) Een uur is vijfenvijftig minuten/Een minuut is zestig seconden
18) Melk bevat alcohol/Wijn bevat alcohol
19) Mensen hebben op de zon gelopen/Mensen hebben op de maan gelopen
20) Sommige mensen drinken thee met zout/Sommige mensen drinken koffie met suiker
21) Olifanten eten soms mensen/Tijgers eten soms mensen
22) Een plant heeft melk nodig om te leven/Een plant heeft water nodig om te leven
23) Een groen licht betekent stop/Een rood licht betekent stop
24) Papier wordt gemaakt van onkruid/Papier wordt gemaakt van bomen
25) Een fiets is een oorlogswapen/Een tank is een oorlogswapen
26) Boeddhisme is een politieke theorie/Hindoeïsme is een religie
27) Spaghetti is een Frans gerecht/Spaghetti is een Italiaans gerecht
28) Een dierenarts kan je helpen als je ziek bent/Een arts kan je helpen als je ziek bent
29) Fietsen is meestal langzamer dan lopen/Fietsen is meestal sneller dan lopen
30) Kinderen zijn nooit bang in het donker/Kinderen zijn vaak bang in het donker
31) Papier is een goed materiaal voor gebouwen/Steen is een goed materiaal voor gebouwen
32) Boekhouden is een kunstvorm/Boekhouden is een beroep
33) Juni is een wintermaand/Juli is een zomermaand
34) Een lamp is een rijtuig/Een auto is een rijtuig
35) Een stoel heeft een motor/Een vrachtwagen heeft een motor
36) Een boot is een soort meubel/Een tafel is een soort meubel

Practice Items:
P1) Een sinaasappel is knapperig/Een wortel is knapperig
P2) Een baksteen is een edelsteen/Een smaragd is een edelsteen
P3) De hoofdstad van Engeland is Manchester/De hoofdstad van de V.S. is Washington

P4) Tom Hanks is een bekend musicus/Tom Hanks is een bekende acteur
APPENDIX C: SPANISH MATERIALS

Appendix C contains the entire Spanish materials set. Sentence pairs are given in the order False, followed by True.

1) Un camello es un tipo de pájaro/Un pelícano es un tipo de pájaro
2) Los tigres tienen alas/Las águilas tienen alas
3) Una col es un tipo de fruta/Una fresa es un tipo de fruta
4) Los frijoles son dulces/Las manzanas son dulces
5) Un gorrión es un mamífero/La vaca es un mamífero
6) Una camisa es una parte del cuerpo/El pie es una parte del cuerpo
7) Los zapatos tienen dedos/Las manos tienen dedos
8) Un mono es un tipo de pez/El atún es un tipo de pez
9) Un taladro es un instrumento musical/El piano es un instrumento musical
10) Un violín es una herramienta/Un martillo es una herramienta
11) La gente usa pantalones en las manos/La gente usa calcetines en los pies
12) Algunas personas tienen girafas como mascotas/Algunas personas tienen perros como mascotas
13) La mayoría de los carros funcionan con jugo de manzana/La mayoría de los carros funcionan con gasolina
14) Dinamarca es un país de Asia/España es un país en Europa
15) Los caballos tienen tres patas/Los caballos tienen cuatro patas
16) Fumar es bueno para la salud/Hacer ejercicio es bueno para la salud
17) Una hora tiene cuarenta y cinco minutos/Un minuto es sesenta segundos
18) La leche contiene alcohol/ El vino tiene alcohol
19) Hay gente que ha caminado en el Sol/Hay gente que ha caminado en la Luna
20) Algunas personas beben té con sal/Algunas personas beben café con azúcar
21) A veces los elefantes comen personas/A veces los tigres comen personas
22) Las plantas necesitan leche para vivir/Las plantas necesitan agua para vivir
23) Un semáforo verde significa 'pare'/Un semáforo rojo significa 'pare'
24) El papel se hace a partir de las malas hierbas/El papel se hace a partir de los árboles
25) Una bicicleta es un arma de guerra /Los tanques son armas de guerra
26) El budismo es una teoría política/ El hinduismo es una religión
27) Los espaguetis son un plato francés/ Los espaguetis son un plato italiano
28) El veterinario te puede ayudar cuando estás enfermo/ Los doctores te pueden ayudar cuando estás enfermo
29) Andar en bicicleta generalmente es más lento que caminar/Andar en bicicleta generalmente es más rápido que caminar
30) A los niños nunca les asusta la oscuridad/A los niños muchas veces les asusta la oscuridad
31) El papel es un buen material de construcción/La piedra es un buen material de construcción
32) La contabilidad es una forma de arte/La contabilidad es una profesión
33) Junio es un mes de invierno/Julio es un mes de verano
34) Una lámpara es un vehículo/Un carro es un vehículo
35) Las sillas tienen motores/Los camiones tienen motores
36) Un barco es un tipo de mueble/Una mesa es un tipo de mueble

Practice Items:

P1) Las naranjas son crujientes/Las zanahorias son crujientes

P2) Un ladrillo es una joya/Una esmeralda es una joya

P3) La capital de Inglaterra es Manchester/La capital de los E.U. es Washington

P4) Tom Hanks es un músico famoso/Tom Hanks es un actor famoso
APPENDIX D: LANGUAGE BACKGROUND QUESTIONNAIRE AND LEXICAL DECISION TESTS

A) The following are the questions contained on the language background questionnaire:

1. How old are you?
2. What do you study (or: what is your job), and in which year?
3. How old were you when you first got in contact with the English language intensively?
4. How did this happen (school, travelling, course, etc.?)
5. How many years of experience with English do you have?
6. How often do you read English literature outside of your study / work?
   - very rarely
   - sometimes
   - regularly
   - very often
7. How often do you read English books / articles for your study / work?
   - very rarely
   - sometimes
   - regularly
   - very often
8. How often do you speak English?
   - very rarely
   - sometimes
   - regularly
   - very often
9. How often do you watch English television or listen to English radio?
   - very rarely
   - sometimes
   - regularly
   - very often
10. In general, how much reading experience do you have with the English language?
    - very little
    - some
    - reasonable
    - very much
11. In general, how much writing experience do you have with the English language?
    - very little
    - some
    - reasonable
    - very much
12. In general, how much speaking experience do you have with the English language?
    - very little
    - some
    - reasonable
    - very much
13. Have you recently (today) been in contact with the English language (text book, course, television/radio, foreign friends, etc.)?
14. Do you have further remarks or comments related to your acquisition or usage of English (courses, stay abroad, English-speaking family, English as your dominant language, etc.)?
15. Have you learned any other foreign language(s), and if so, how often do you use it/them?
   language 1(if applicable):
   - very rarely
   - sometimes
   - regularly
   - very often
   Do you use this language more often than English? yes no
B) The following is the list of words and nonwords comprising the offline lexical decision test; the first three items were considered practice, resulting in a 60 item list of 40 words and 20 orthographically legal nonwords. The words are followed by their lexical frequency (Francis & Kucera, 1986) in parentheses:

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platery</td>
<td></td>
</tr>
<tr>
<td>Denial</td>
<td></td>
</tr>
<tr>
<td>Generic</td>
<td></td>
</tr>
<tr>
<td>Mensible</td>
<td></td>
</tr>
<tr>
<td>Scornful</td>
<td></td>
</tr>
<tr>
<td>Stoutly</td>
<td></td>
</tr>
<tr>
<td>Ablaze</td>
<td></td>
</tr>
<tr>
<td>Kermshaw</td>
<td></td>
</tr>
<tr>
<td>Moonlit</td>
<td></td>
</tr>
<tr>
<td>Lofty</td>
<td></td>
</tr>
<tr>
<td>Hurricane</td>
<td></td>
</tr>
<tr>
<td>Flaw</td>
<td></td>
</tr>
<tr>
<td>Alberation</td>
<td></td>
</tr>
<tr>
<td>Unkempt</td>
<td></td>
</tr>
<tr>
<td>Breeding</td>
<td></td>
</tr>
<tr>
<td>Festivity</td>
<td></td>
</tr>
<tr>
<td>Screech</td>
<td></td>
</tr>
<tr>
<td>Savory</td>
<td></td>
</tr>
<tr>
<td>Paludate</td>
<td></td>
</tr>
<tr>
<td>Shin</td>
<td></td>
</tr>
<tr>
<td>Fluid</td>
<td></td>
</tr>
</tbody>
</table>

Magrity  
Allied 0  
Slain 1  
Recipient 9  
Exprate  
Eloquence 2  
Cleanliness 6  
Dispatch 8  
Rebondicate  
Ingenious  
Bewitch 3  
Skave  
Plaintively 2  
Kilp  
Interfate  
Hasty 5  
Lengthy 11  
Fray 1  
Crumer  
Upkeep 6  
Majestic 9  

Nourishment 1  
Abery  
Proom  
Turmoil 12  
Carbohydrate 1  
Scholar 8  
Turtle 4  
Fellick  
Distription  
Cylinder 3  
Censorship 4  
Celestial 4  
Rascal 1  
Purrage  
Pulsh  
Muddy 2  
Quirty  
Poudor  
Listless 1  
Wrought 1
REFERENCES


