The Effects of Language Background and Foreign Accent on Listening Comprehension

Sita Carraturo
The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Psycholinguistics and Neurolinguistics Commons

Recommended Citation
Carraturo, Sita, "The Effects of Language Background and Foreign Accent on Listening Comprehension" (2019). CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/3501

This Thesis is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE EFFECTS OF LANGUAGE BACKGROUND AND
FOREIGN ACCENT ON LISTENING COMPREHENSION

by

SITA CARRATURO

A master’s thesis submitted to the Graduate Faculty in Liberal Studies in partial fulfillment of
the requirements for the degree of Master of Arts, The City University of New York

2019
The Effects of Language Background and Foreign Accent on Listening Comprehension

by

Sita Carraturo

This manuscript has been read and accepted for the Graduate Faculty in Liberal Studies in satisfaction of the thesis requirement for the degree of Master of Arts.

Date

Gita Martohardjono

Thesis Advisor

Date

Gita Martohardjono

Executive Officer

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

The Effects of Language Background and Foreign Accent on Listening Comprehension

by

Sita Carraturo

Advisor: Gita Martohardjono

The act of listening to a linguistic signal is an involved process, and, rarely occurs in absolute silence. A person trying to listen and comprehend speech is likely in an environment that has some sort of additional noise: white noise from a fan, passing traffic, construction, or just other talkers. Each of these additional auditory signals creates an unfavorable environment for the listener who is trying to capture the target signal. Research has been able to quantify and describe the effects of noise on the comprehension of linguistic signals, and has also shown that that bilinguals and monolinguals — though their performance is indistinguishable in quiet conditions — are known to be differentially affected by noise: bilinguals perform significantly worse in adverse listening conditions when tasked with comprehend a linguistic signal. What is yet to be established is how a signal with intrinsic, phonological variation differentially affects monolinguals and bilinguals.

This study is a small-scale pilot that investigates this question: what bearing does bilingualism have on the comprehension of foreign-accented speech in quiet and in noise? Stimuli include sentences spoken in English, with five different accents: Neutral English (the English typical of the NYC area), Latin American Spanish English, Mandarin English, Italian English, and
Indian English. A true-false verification task is used to assess the participants’ comprehension of the sentences, which are auditorily delivered such that no two sentences with the same accent are heard consecutively. All five accents are heard in both quiet and in noise, in two separate blocks. Accuracy and reaction times are analyzed.
I would like to thank my advisor, Dr. Martohardjono, for her guidance in this project. Special thanks, also, to Dr. Loraine K. Obler, whose instruction, example, and support were all invaluable to me during my years at the Graduate Center. To the following friends and colleagues, thank you so much for your advice and undying support through the tough times: Iris Strangmann, Taryn Malcolm, Katarina Antolovic, Jungna Kim.
# TABLE OF CONTENTS

List of Tables vii

Chapter 1: Introduction 1

1.1 The Auditory Signal and Listening Comprehension 1

1.2 Signal-Extrinsic Factors: SNR 1

1.3 Listener-Intrinsic Factors 2

1.4 Signal-Intrinsic Factors 3

1.5 Intrinsic and Extrinsic Degradation 4

1.6 Bilinguals and Foreign Accented Speech Perception 6

Chapter 2: Current Study 7

2.1 Motivation and Hypotheses 7

2.2 Stimuli 8

2.3 Participants 11

2.4 Procedure 12

Chapter 3: Results and Discussion 13

3.1 Results 13

3.2 Discussion 13

Bibliography 15
LIST OF TABLES

Table 1: Speaker Information 10
Table 2: Participant Information 11
CHAPTER 1: INTRODUCTION

1.1 The Auditory Signal and Listening Comprehension

Language comprehension is a complex process that includes parsing, lexical retrieval, interpretation of pragmatics, and more. All of this, however, is based on the linguistic signal received; in spoken language, this would be an auditory speech signal.

In pristine speaking and listening conditions, the success of language comprehension and the time it takes to be completed are most likely to be modulated by linguistic factors such as the complexity of the syntax or the frequencies of the lexical items. Pristine speaking and listening conditions, however, are rare. The intelligibility of an auditory signal can be degraded due to a number of factors: speaker-intrinsic factors (e.g., speech impediment, accent), listener-intrinsic factors (e.g., hearing impairment), and/or external factors (e.g., environmental noise). Given this range of variables that can affect the speaking and listening conditions, it is likely that the intelligibility of the speech signal will be degraded in some way, and yet clarity in communication is a universally desirable trait. As such, a large number of studies have sought to measure the impact of these factors in various combinations.

1.2 Signal-Extrinsic Factors: SNR

Possibly the most common and unavoidable form of signal degradation is environmental noise. Environmental or background noise can come in a number of forms: a loud air conditioner, other people speaking within the same vicinity, static over a telephone, or an airplane passing overhead, just to name a few. When these types of noises present themselves at the same time as a speech signal, they effectively compete for the listener’s attention, and in doing so, negatively affect the listener’s ability to comprehend the speech signal.
A large body of research has been done to quantify and describe the effects of noise on the comprehension of linguistic signals. The signal-to-noise ratio (SNR) is a crucial predictor of the impact of the noise: the louder the noise as compared to the target signal, the poorer the comprehension of said signal (e.g., Adank et al., 2009; Shi, 2010; Tabri et al., 2011).

These findings hold insofar as listener variables are accounted for: studies have investigated how adverse listening conditions, in combination with certain listener-intrinsic variables, reveal distinctions between different groups of people. These studies are discussed in the following section.

1.3 Listener-Intrinsic Factors

Aside from hearing ability, age is a well-known factor that affect a person’s ability to comprehend an auditory speech signal (e.g., Ferguson et al., 2010; Adank & Janse, 2010). Recent studies have found that that bilinguals and monolinguals — though their performance is indistinguishable in quiet listening conditions — are differentially affected by noise: bilinguals perform significantly worse when tasked with processing a linguistic signal in noise.

For example, Krizman et al. (2016) studied a group of high school students (31 English monolinguals and 25 Spanish-English bilinguals), and found that monolinguals performed better than bilinguals in both word-in-noise and sentence-in-noise tasks, though only the latter yielded statistically significant results. This same group of bilinguals, however, performed significantly better than their monolingual counterparts in a test of tone-in-noise (non-linguistic). Though Krizman and colleagues did not find significant results for the word-in-noise task, both Tabri et al. (2011) and Rogers et al. (2006) did find that the monolinguals significantly outperformed the bilinguals in single word tasks.
It is important to note that Krizman et al. (2016), Tabri et al. (2011), and Rogers et al. (2006) tested groups of bilinguals who were highly proficient in English, and each group of bilinguals was homogenous (acquisition of languages prior to 6 years of age; prior to 5 years in Krizman et al.). Earlier comparisons of bilinguals and monolinguals in speech-in-noise tasks often focused on bilinguals who were exposed to English at varying ages of exposure or acquisitions of the L2 (Mayo et al. 1997; Meador et al., 2000). Findings have shown that proficiency, length of period of immersion, and age of fluency all contribute to the outcomes (Shi, 2012).

1.4 Signal-Intrinsic Factors

While the aforementioned studies have demonstrated that these groups perform differently in the presence of background noise (a signal-extrinsic factor), little research has been done to establish whether monolinguals and bilinguals perform differently in the presence of a deterioration that is intrinsic to the signal itself.

The type of intrinsic deterioration that is of interest to the current study is foreign accent. Rather than “deterioration”, variation is a term that better encapsulates foreign accent. Munro & Derwing (1995a) found that nonnative speech does require more processing time than native-accented speech, and earlier studies in the literature showed a negative bias that native speaker listeners held against non-native speech (e.g., Anisfeld et al., 1962; Kalin & Rayko, 1978). However, while accents can be deleterious to communication, they are not always. There are two relevant measures regarding foreign accent: accentedness and intelligibility. According to another study by Munro and Derwing (1995b), these two measures are partially independent of
one another. That is, a speaker whose speech is considered to be highly accented can also be found to be highly – if not entirely – intelligible.

To date, studies of the effect of foreign accent on comprehension seem to be done on two populations: monolinguals, and late L2 learners. Munro and Derwing (1995a, 1995b) tested both the intelligibility and processing costs of foreign accented speech in monolingual listeners. Processing cost was evaluated by the response latency in a true/false sentence verification task in which the stimuli were spoken by either native Mandarin speakers or native English speakers (50/50 mix). Listeners were native speakers of English. After listeners indicated whether the sentence they just heard was true or false, they then wrote out the sentence that they heard, thereby providing the researches with a measure of intelligibility. In this study, both response time and intelligibility showed significant results: sentences spoken by the L1-Mandarin speakers resulted in longer response latencies than those of the L1-English speakers, and significantly fewer L1-Mandarin speaker sentences were correctly transcribed (Munro & Derwing, 1995b).

Adank et al. (2009) found that among monolingual listeners, processing was slower for unfamiliar regional accents than that for familiar ones, and slowest for unfamiliar foreign accents. Rogers et al. (2004) also found that the intelligibility of foreign accented speech (English sentences, spoken by native Mandarin speaker) was significantly more affected by noise than was non-accented speech, but this again was a study done only with monolingual listeners.

1.5 Intrinsic and Extrinsic Degradation

Before moving on to bilinguals and foreign accented speech perception, it warrants taking a look at studies have been done to investigate the interaction between accented speech and noise. Wilson & Spaulding (2010), and Rogers et al. (2004) both looked at how different
SNRs could interact with accented English by speakers of varying proficiency levels, resulting in four experimental conditions. Both studies found that speaker intelligibility was highest for native speaker, lower for high proficiency speakers, and lowest for low proficiency speakers. Crucially, both studies also found that a moderate amount of masking noise (+10 db SNR) caused a significant decrease in the intelligibility of high-proficiency speakers (though no change for native speakers). In addition to these intelligibility scores, Wilson & Spaulding, (2010) also looked at processing time and found that listeners’ reaction times were shortest for native proficiency speakers, and longest for low-proficiency speakers.

These findings from Rogers et al. (2004) and Wilson & Spaulding (2010) contradict those from other studies such as Ferguson et al. (2010) and Lane (1963) who found that comprehension in noise was not affected by speaker accent. These disparities, however, are likely due to differences in methodology. The findings from Rogers and colleagues as well as from Wilson & Spaulding both demonstrate clearly that noise and accent only interact in either of two cases: a low SNR (+5 db; Rogers et al., 2004), or a high intelligibility rating of the speaker in quiet (high proficiency speakers). Low-proficiency speakers combined with high SNRs will not yield a significant decrease in comprehension, as compared to the same speaker in quiet.

Together, these studies show that the obstacle presented by background noise is distinct from the one presented by accent. In fact, Ferguson et al. (2010) found that older adults were not affected by foreign accent any more than younger adults were. Mattys et al. (2012) points out two main divisions between accent and background noise: the first is the obvious: that foreign accent is an intrinsic signal degradation while the degradation from noise is extrinsic. Secondly,
accent is a patterned degradation: the phonological variation is (almost by definition) consistent and predictable.

This predictability is possibly what allows listeners to adapt to foreign accented speech. Clarke & Garrett (2004) demonstrated that listeners can adapt to a foreign accent within a single minute of exposure to that accent. Weber et al. (2014) went further and tested two participant groups (monolinguals and bilinguals), and two types of accents (genuine and arbitrary). In this study, the investigators found that not only did native English monolinguals adapt to English words with an Italian accent, but so did Dutch-English bilinguals: they were able to accommodate the Italian-accented words of their L2. Most notable among Weber et al.’s findings is that the Dutch-English bilinguals were able to accommodate an arbitrary accented Dutch as well as arbitrary accented English, while the monolinguals failed to do so. The authors claim that this finding likely indicates that the phonetic-to-lexical mapping is a flexible process among bilinguals. This is supported as well by the findings of Samuel & Larraza 2015.

1.6 Bilinguals and Foreign Accented Speech Perception

More recent studies have continued to look at the effects of foreign accented speech in various ways, studying its relationship to intelligibility (Gittleman & Van Engen, 2018), individual differences in its perception (McLaughlin et al., 2018), as well as the neural substrates related to its processing (Adank et al., 2012). However, what is missing in the literature is an understanding of how bilinguals, specifically, cope with foreign accented speech. That being said, it is not entirely unexplored.

Larraza et al. (2016) studied how Spanish-Basque and French-Basque bilinguals performed on phoneme discrimination task in order to understand the effects of L1 and age of
acquisition (AoA) on accented-speech perception. Their findings showed that AoA contributes significantly to the costs and accuracies measured: the simultaneous bilinguals across both L1s performed the best. Additional studies of foreign accent and bilinguals have focused on late L2 learners who themselves speak with an accent, and found that the accent of the listener directly affects their performance on accented speech comprehension (e.g., Bent & Bradlow, 2003).

What is of interest for the current study, however, is how the life-long exposure to a second phonology, i.e., simultaneous bilingualism, modulates the experience of processing sentence-level foreign accented speech. This, to the knowledge of the author, remains unstudied.

CHAPTER 2: CURRENT STUDY

2.1 Motivation and Hypotheses

As discussed in the previous sections, findings from other studies demonstrate that bilinguals perform significantly worse than monolinguals in speech in noise comprehension tasks (e.g., Lucks Mendel & Widner, 2016). While this effect of external signal degradation is well accounted for, the literature lacks in studies that have looked at how bilingual performance is affected by signal-intrinsic deterioration, like foreign accent, particularly at the sentence level. There is reason to believe that bilinguals may have a more flexible in their mapping from phonological to lexical form (Samuel & Larraza, 2015). This pilot-study aims to investigate just that. Specifically, the questions motivating this study are:

1) How does a speaker’s foreign accent affect speech comprehension among bilinguals?
2) Does the relationship between a speaker’s L1 and a listener’s L1 have an impact on the listener’s foreign accented speech comprehension?
3) What are the compounded effects of masking noise and foreign accent on speech
comprehension?

In the current study (a small-scale pilot study), participants performed a sentence-verification task. They heard English sentences, each of which was presented in one of five accents: four foreign accents and one unmarked “home accent” (i.e., the accent that is common to the area they live in, Duffy 2013). The participants were asked to determine whether each sentence was true or false. Two measurements were recorded: accuracy, and response time. The participants constituted two groups: Spanish-English early bilinguals (English-dominant), and English monolinguals. The stimuli, too, constituted two types: half of the sentences were presented in quiet, the other half were overlain with speech-modulated noise.

Because all of the stimuli were English sentences, the participants were screened to ensure they were all English-dominant. Bilingual participants were screened to ensure that they had begun to acquire both languages before the age of five, and spoke English on a daily basis, while it was ascertained that the monolinguals — who would have studied at least one language other than English during their school years — did not speak a second language on a regular basis.

Based on evidence from prior studies, the hypotheses are as follows:

1) Across groups:

   a. Bilinguals would outperform monolinguals when the sentences presented with a foreign accent in quiet.

   b. Monolinguals would outperform bilinguals in the noisy condition

2) Within-group:

   a. Monolinguals would perform worst on foreign-accented sentences in noise
b. Bilinguals would demonstrate an advantage on Spanish-accented sentences in quiet and in noise

As mentioned earlier, this study is a pilot study aimed at gathering evidence of these hypotheses for a future, larger-scaled study.

2.2 Stimuli

There were 144 English sentences prepared: half of them true statements (e.g., Many children go to school), half of them false (e.g., People have twelve fingers and toes). Of these sentences, 33 originate from the published stimuli in Munro & Derwing’s study (1995b), and the rest were modeled after those sentences. In the end, 65 true and 65 false sentences were chosen to be presented; each participant had an additional 3 practice trials. All participants heard the same sentences spoken by the same speakers.

Because the listeners (the participants) would all be native English speakers, the stimuli were not controlled for lexical frequency; however, sentence length (based on syllable count) was controlled for (avg: 8.22 syllables per sentence; range: 7-10 syllables).

The stimuli were then recorded by female speakers of English with various L1s, such that each stimulus sentence was spoken and recorded in each of the following accents: (1) Mandarin, (2) Colombian Spanish, (3) Italian, (4) South Indian/British, and (5) NYC Regional. The four foreign accents were chosen based on the availability of speakers, whereas the fifth, ‘neutral’ accent was chosen to serve as a measurement of baseline performance. Following is a description of each of the speakers.

2.2.1 Speaker Descriptions
The first speaker is 29 years old, and born and raised in China. This speaker’s L1 is Mandarin and she began studying English in elementary school. According to the speaker, she did not actively speak English until she entered university.

The second speaker is 30 years old, from Colombia. Her L1 is Spanish. She began to learn English when she moved to the USA ten years ago. Though she was living in the US, she lived in a predominantly Spanish-speaking area of Queens, New York, which limited her exposure to and need to speak English.

The third speaker is 30 years old, and was born and raised in Italy; her L1 is Italian. She moved from Italy to the USA three years prior to her involvement in this project. Though she had English education throughout her grade-school years, she claims that she really only began to acquire English when she studied it for three years in college.

The fourth speaker is 32 years old, born and raised in the Maldives and Southern India. She moved to the USA at the age of 24, but had begun studying English from the age of 3 when she entered the school system. Her L1 is Tamil, though she was not educated in this language. Having gone to school in India, the target-English in the classroom was not American English, but rather British and Indian Englishes. Her resulting accent is evidently a blend of British English and Tamil.

The fifth speaker is 29 years old, born and raised in the New York City metropolitan area. Though multilingual, English was one of this speaker’s first two languages; it was acquired from birth (spoken in the household) and served as her primary language of instruction throughout school. Her accent is typical of speakers from the NYC area, and is therefore considered unmarked for this study and participant population.
Table 1: Speaker Information

<table>
<thead>
<tr>
<th>L1</th>
<th>Age</th>
<th>Age of English Acquisition</th>
<th>Years living in English-speaking country</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Mandarin</td>
<td>29</td>
<td>&lt; 9</td>
<td>1.5</td>
</tr>
<tr>
<td>(2) Spanish</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>(3) Italian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Tamil</td>
<td>32</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>(5) English</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
</tbody>
</table>

2.2.2 Stimuli Preparation

Stimuli were recorded in a sound-attenuated booth for maximum quality. Speakers were instructed to read the sentences in a clear voice; in the event of an unnatural pause, error, or unsuitable volume, the speakers were asked to repeat the sentence or given additional instruction as needed.

Once all sentences were recorded, each sentence was normalized in Praat at 60 dB. These files were then duplicated such that one set of them were left in the clear and normalized condition, while the second half were overlain with speech-modulated noise at an SNR of -3.

2.3 Participants

Participants were grouped into two language-based categories: monolingual English speakers, and English-Spanish bilinguals (English dominant speakers who were early Spanish-English bilinguals). There were six participants total, three in each group. All participants were 25-34 years old (average age: 30). Participants had similar levels of education, as all were at varying levels of doctoral study.

All participants completed a questionnaire, providing information regarding their language backgrounds. No participants indicated having any auditory impairments or
neurological diseases. Individual participant data is as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex</th>
<th>Age</th>
<th>Lang Study</th>
<th>Spanish Literacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Monolingual</td>
<td>F</td>
<td>34</td>
<td>Spanish, Japanese</td>
<td>None</td>
</tr>
<tr>
<td>(2) Bilingual</td>
<td>M</td>
<td>29</td>
<td>Mandarin</td>
<td>Reading</td>
</tr>
<tr>
<td>(3) Bilingual</td>
<td>M</td>
<td>33</td>
<td>Spanish, French</td>
<td>Reading &amp; Writing</td>
</tr>
<tr>
<td>(4) Monolingual</td>
<td>M</td>
<td>34</td>
<td>Spanish, Japanese</td>
<td>None</td>
</tr>
<tr>
<td>(5) Monolingual</td>
<td>M</td>
<td>25</td>
<td>Spanish</td>
<td>Reading &amp; Writing</td>
</tr>
<tr>
<td>(6) Bilingual</td>
<td>F</td>
<td>25</td>
<td>Spanish</td>
<td>Reading &amp; Writing</td>
</tr>
</tbody>
</table>

2.3.1 Listener Language Backgrounds

As noted in the table above, all participants reported having studied at least one other language for at least one year in school. The maximum number of years reported of foreign language study is six years; this was reported by two monolingual speakers.

The categorization of participants into the two groups was not based on second language ability, but rather on the criteria that the bilinguals would have acquired both English and Spanish prior to the age of 3, and that the participants’ daily activities either did or did not involve communicating in a second language (Spanish). No monolinguals report speaking a second language, and no bilinguals report speaking a third language.

2.4 Procedure

Using EPrime software, the stimuli were organized into two blocks: quiet and noise. Three of the participants heard the quiet block first, while the other three heard the noise block.
first. The order of the stimuli within each block remained the same. This order was pseudo-randomized to ensure that no two sentences of the same accent were presented consecutively, and that the truth value of consecutive sentences did not present in any particular pattern.

Participants were seated in a sound-attenuated booth facing a computer. Once seated comfortably, the participants were instructed, both orally and in written form on the computer screen, to listen to the speech stimuli (delivered through speakers) and to decide as quickly and as accurately as possible whether the sentence they just heard was true or false. The participants marked their response by pressing one of two keys with their right hand. Once they entered their response, the software cued the next stimuli, which would be presented after a two-second delay. Following the initial instructions, the participants practiced with three sentences and were then prompted to ask questions if they had any. The experiment was self-paced and measured both accuracy and response time.

CHAPTER 3: RESULTS AND DISCUSSION

3.1 Results

3.1.1 Accuracy

A multinomial logistic regression analysis was performed within each of the two conditions (noise, quiet). Accent information was coded binarily: 0 for the home accent items, 1 for any of the foreign accent items. Effect of group and language were analyzed, but only group yielded a significant effect (p = .000) in the noise condition; there were no significant effects in the quiet condition.

3.1.2 Reaction Time
Prior to analysis, all reaction time data for inaccurate responses was eliminated, as were extreme outliers. A linear mixed model analysis was performed to analyze the remaining reaction time data in each condition. Again, no effect of language was found. Group, however, had a significant effect in both conditions: noise (p = .001) and quiet (p=.000).

3.2 Discussion

The results from these data do not indicate that accented speech affects the listening comprehension as measured by accuracy and response time, regardless of whether the listener is monolingual or bilingual. However, as earlier studies have noted that accented speech does affect monolingual listening comprehension, these data cannot be considered to reliably represent the behaviors of these populations in these environments. Certainly, a larger number of participants may have resulted in more telling results, but additional limitations to the study include the inconsistencies among the speakers, and the disproportionate variable to control ratio in the experimental design.

That group yielded a significant effect in noise/accuracy analysis is consistent with studies in the literature that have demonstrated bilinguals performing more poorly than monolinguals in speech-in-noise comprehension tasks.

3.2.1 Speaker Limitations

While each speaker varied in some way from the others, from among the five speakers, two in particular were perceptibly distinctive. The speaker with a British Indian accent spoke more animatedly than the others, with a cadence and prosody one might use to read to young
children. Moreover, relative to the listener’s languages, this speaker’s speech was indicative of two non-native phonologies, as opposed to one.

The second speaker whose speech was particularly different was the Colombian Spanish speaker. This speaker’s speech was less clear than the other speakers’ due to characteristics of her speech that could not be accounted for by phonology alone. For example, one participant indicated that he felt this speaker “mumbled”.

### 3.2.2 Experimental Limitations

One flaw in the design of this study is the proportion of foreign accented items to home-accented items: 80% of the stimuli were foreign accented. The stimuli were divided equally such that each of the five accents accounted for 20% of the stimuli, but four of the five speakers had foreign accents. Combined with the small N, there is insufficient power for reliable analysis
Bibliography


