

Second Language Comprehensibility Revisited: Investigating the Effects of Learner Background

Dustin Crowther

A Thesis
In
The Department
of
Education

Presented in Partial Fulfillment of the Requirements
for the Degree of Masters of Arts (Applied Linguistics) at
Concordia University
Montréal, Québec, Canada

April 2014

© Dustin Crowther, 2014

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Dustin Crowther

Entitled: L2 Comprehensibility Revisited: Investigating the Effects of Learner Background
and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

Complies with the regulations of the University and meets the accepted standards with respect to
originality and quality.

Signed by the final Examining Committee:

_____ (Chair)

Walcir Cardoso

_____ (Examiner)

Kim McDonough

_____ (Examiner)

Sara Kennedy

_____ (Examiner)

Kazuya Saito

_____ (Supervisor)

Pavel Trofimovich

Approved by:

Graduate Program Director

Dean of Faculty

Date: _____ 2014

ABSTRACT

Second Language Comprehensibility Revisited: Investigating the Effects of Learner Background

Dustin Crowther

The current study investigated first language (L1) effects on listener judgment of comprehensibility and accentedness in second language (L2) speech. The participants were 60 university-level adult speakers of English from four L1 backgrounds (Chinese, Romance, Hindi, Farsi), with 15 speakers per group, performing a picture narrative task. Ten native English listeners used continuous sliding scales to evaluate the speakers' audio recordings for comprehensibility, accentedness, as well as 10 linguistic variables drawn from the domains of pronunciation, fluency, lexis, grammar, and discourse. While comprehensibility was associated with several linguistic variables (segmentals, prosody, fluency, lexis, grammar), accentedness was primarily linked to pronunciation (segmentals, word stress, intonation). The relative strength of these associations also varied as a function of the speakers' L1, especially for comprehensibility, with Chinese speakers influenced chiefly by pronunciation variables (specifically segmental errors), Hindi speakers by lexicogrammar variables, Romance speakers by variables spanning both pronunciation and lexicogrammar domains, and Farsi speakers showing no strong association with any linguistic variable. These results overall suggest that speakers' L1 plays an important role in listener judgments of L2 comprehensibility and that instructors aiming to promote L2 speakers' communicative success may need to expand their teaching targets beyond segmentals to include prosody-, fluency-, and lexicogrammar-based targets.

Keywords: Comprehensibility; accentedness; L1 Influence; pronunciation learning

Acknowledgment

I would like to offer my most sincere gratitude to my supervisor, Dr. Pavel Trofimovich, for his support and guidance throughout this process. His patience and enthusiasm were not only motivating, but created an environment that was a pleasure to work in. As well, I would like to thank my readers, Drs. Kim McDonough, Sara Kennedy, and Kazuya Saito, for their feedback, support, and time. Finally, I would like to express my appreciation for my family and friends, whose love and support during the last two years have helped me to maintain my sanity throughout this busy and stressful time.

Contribution of Authors

As first author for the included manuscript, Dustin Crowther was primarily responsible for all aspects of this study, including conceptualization, the development of materials, participant testing, data analysis, and interpretation, with the support of Dr. Pavel Trofimovich, who provided guidance and feedback throughout the various stages. Dustin Crowther was responsible for the development of substantive drafts and redrafts of the manuscript, with Dr. Trofimovich supplying extensive feedback on organization and content, as well as rewording some passages. The majority of the manuscript's conception and content was developed by Dustin Crowther, which is reflected in his status as first author.

Further, Drs. Kazuya Saito (Waseda Univeristy) and Talia Isaacs (University of Bristol) both provided valuable contributions to the development and completion of this study. Dr. Saito was responsible for the development of the instrument (rating scale) that was subsequently modified for use within this study while Dr. Isaacs contributed to the corpus of speech data used for rating and analysis during this study. Their contributions are thus recognized through a co-authorship of the included manuscript.

Table of Contents

List of Tables	vii
Chapter One	1
Manuscript Literature Review	5
Disentangling Comprehensibility from Accent	6
L1 Effects on L2 Comprehensibility and Accentedness	9
The Current Study	11
Method Participants	12
Speakers	12
Raters	14
Materials	14
Speech Rating	15
Rated Categories	16
Accent and comprehensibility	16
Phonology and fluency	17
Lexicon, grammar, and discourse structure	18
Understanding and use of rated categories	19
Results Rater Consistency	20
Comprehensibility and Accent	20
Ratings of Linguistic Dimensions	22
L1 Background and Linguistics Dimensions on	
Comprehensibility and Accentedness	24
Discussion	26
Pronunciation Variables and Accentedness	27
Pronunciation Variables and Comprehensibility	28
Limitations and Future Research	31
Conclusion	34
Chapter Three	36
References	42
Appendix	52

List of Tables

Table 1	Means and Standard Deviations for Participant Background Characteristics	13
Table 2	Means and Standard Deviations for Accent and Comprehensibility Ratings by L1 Group (Based on a Continuous 1000-Point Sliding Scale) .	21
Table 3	Summary of a Two-Factor Solution Based on a Principal Component Analysis of the Ten Rated Linguistic Variables	23
Table 4	Results of Multiple Regression Analyses Using the Factors of Pronunciation and Lexicogrammar as Predictors of Accent and Comprehensibility	23
Table 5	Pearson Correlations Between the Pronunciation and Lexicogrammar Factors and Accent and Comprehensibility by L1 Group	24
Table 6	Pearson Correlations Between the 10 Rated Linguistic Categories and Accent and Comprehensibility by L1 Groups	25

Chapter One

In conceptualizing second language (L2) pronunciation learning, two principles have been put forth, defined by Levis (2005) as the *nativeness principle* and the *intelligibility principle*. These two principles have been the focus of a large body of L2 pronunciation literature, with both sides receiving extensive attention (see Munro & Derwing, 2011, for an overview of existing research). The nativeness principle targets native-like, non-accented L2 pronunciation as a desirable learning and teaching goal, an idea shared by many L2 speakers and their native-speaking interlocutors alike. In a survey of adult immigrants studying English in Canada, Derwing (2003) found that 95% desired to sound like a native speaker of English. Of the 100 participants, 55 felt their communicative breakdowns were due to pronunciation-related difficulties, and 59% of the visible minority learners within the group felt that non-accented pronunciation would garner them more respect from native-speaking Canadian citizens. A similar survey of university students from Tokumoto and Shibata (2011) revealed that 68% of 50 Japanese participants and 59% of 46 Korean participants studying in their home country thought that their accented English limited their ability to communicate effectively. These concerns of the L2 speakers are not unfounded. Munro (2003) identified stereotyping, harassment, and even occupational loss as negative ramifications of speaking accented English. Such examples included landlords informing potential, accented tenants that vacant apartments were unavailable, coworkers mimicking accents as a means of ridicule, and administrative personnel labeling a substitute teacher's work file with a note stating he "did not speak English".

In contrast, the *intelligibility principle* emphasizes L2 speakers' ability to be understood over nativelikeness, which is possible despite the presence of foreign accent (e.g., Derwing & Munro, 1997; Derwing & Munro, 2009). Key to the intelligibility principle are the constructs of

intelligibility and comprehensibility. As discussed above, accent refers to a listener's perception of how nativelike an L2 speaker's utterance sounds (Munro & Derwing, 1999). In contrast, intelligibility and comprehensibility are used to capture the listener's ability to understand an L2 speaker. While related, these two constructs reveal different information. Operationalized through scalar ratings, comprehensibility is a measurement of listeners' perceived ease or difficulty of understanding L2 speech, while intelligibility is intended to measure listeners' actual understanding, often through the use of orthographic transcriptions (Munro & Derwing, 1999). In relating these three concepts to the intelligibility principle, Derwing and Munro (1997) found that English-speaking listeners tended to judge accent more harshly than comprehensibility, and comprehensibility more harshly than intelligibility. L2 speakers can actually be judged to possess heavy accents while at the same time being highly comprehensible and intelligible, although speakers who are unintelligible are always rated as being heavily accented (Derwing & Munro, 2009). Based on these research findings, then, the focus on understanding that defines the intelligibility principle is well founded, although this focus clearly clashes with the beliefs of those supporting the nativeness principle.

While individual learner needs obviously cannot be ignored, it must also be understood that there are factors beyond learners' control that may affect their ability to attain nativelike pronunciation, the seemingly most important of which is age of learning (see Major, 2001, and Ortega, 2008, for recent overviews). The vast majority of existing research indicates that the younger the learner, the more likely he or she will acquire nativelike pronunciation (Major, 2001), though starting young is hardly a guarantee of nativelike acquisition (e.g., Flege, Munro, & MacKay, 1995). For older learners, the possession of a foreign accent is seen as normal, often times unavoidable (e.g., MacKay, Flege, & Imai, 2006; Major, 2001; Munro, Derwing, &

Morton, 2006). In the rare instances where older learners do demonstrate nativelike or near nativelike pronunciation (e.g., Bongaerts, 1999; Bongaerts, Mennen, & Van der Slik, 2000; Moyer, 1999) there are usually contributing circumstances, such as amount of exposure, motivation, and type of training, that are unlikely to apply consistently across learners. The prevalence of such difficulties in acquiring nativelike pronunciation further adds support to Levis' (2005) intelligibility principle as not only the target that L2 speakers should aim for, but as the focus of L2 pronunciation training as well.

In terms of L2 pronunciation learning, early research focused primarily on the acquisition of individual vowel or consonant phonemes, and has more recently begun to target aspects of prosody and fluency, such as intonation, rhythm, connected speech, and voice quality settings (Celce-Murcia, Brinton, & Goodwin, 2010). This is reflected in the typical instructional emphasis on segmentals in L2 classrooms, as teachers seem to lack general knowledge and confidence in teaching other, non-segmental speech factors (Derwing, 2003; Foote, Holtby, & Derwing, 2011). This lack of knowledge and confidence likely reflects a gap between research findings and their practical application to the L2 classroom (Celce-Murcia et al., 2005; Derwing & Munro, 2005). And at least one reason for this gap is that both researchers and teachers are still unclear as to which linguistic factors in learner speech are linked with nativelikeness and which contribute to making L2 speech intelligible and comprehensible.

According to existing research, listeners' perception of L2 speech is related to phonological variables such as segmental accuracy (Derwing, Munro, & Wiebe, 1998), word stress (Field, 2005), sentence stress (Hahn, 2004), speech rate (e.g., pausing, articulation rate) (Munro & Derwing, 2001; Trofimovich & Baker, 2006) and pitch range, stress, and pause or syllable length (Kang, Rubin, & Pickering, 2010; Tajima, Port, & Dalby, 1997; Winters &

O'Brien, 2013). Beyond these measures, associations have also been found to exist between intelligibility and grammatical accuracy (Munro & Derwing, 1999) and lexical choice (Fayer & Krasinski, 1987), two factors that are rarely considered when approaching pronunciation training. While the above research provides some insights into which factors affect the perception of L2 speech, they do not give any estimate of the relative importance of these factors to learners' communicative success. Typically, these studies consider only a few of these factors in isolation from each other, and typically only in relation to either accentedness or understanding.

While a series of recent studies (e.g., Saito, Trofimovich, & Isaacs, submitted a, b; Trofimovich & Isaacs, 2012) have begun investigating the relative weight of multiple linguistic factors to measures of understanding and nativelikeness of L2 speech, there remains a great many other variables (e.g., L1 background, task effect, listener experience) that must be considered before any solid recommendations can be made on how this research can be implemented into a classroom approach to L2 acquisition. The goal of the current study is therefore to highlight one such variable, L1 background, and investigate the role it plays in how native speakers of English evaluate the spoken utterances of L2 learners in terms of nativelikeness and understanding.

Second language comprehensibility revisited: Investigating the effects of learner background

With languages such as English, Spanish, Arabic, or Chinese becoming ever more prominent in international trade, education, and popular culture, particularly in communication among non-native speakers, understanding various subcomponents of second language (L2) speaking ability emerges as an important goal for both language researchers and teachers. L2 pronunciation (one aspect of speaking ability) has typically been discussed with reference to two broad constructs, namely, understanding and nativelikeness. The construct of understanding embraces various aspects of speakers' ability to make themselves understood. Following common research and assessment practice, understanding is often measured as comprehensibility or listeners' perception of how easy or difficult it is for them to understand L2 speech, rated on 7- or 9-point scales (Derwing & Munro, 2009). The construct of nativelikeness, which broadly refers to speakers' ability to approximate speech patterns of the target-language community, is usually operationalized as a listener-based rating of accentedness, also using 7- or 9-point scales (Derwing & Munro, 2009).

While it is often the case that many teachers and students view nativelike pronunciation as the ultimate goal of L2 learning (Derwing, 2003; Levis, 2005; Tokumoto & Shibata, 2011), researchers have consistently underscored comprehensibility as a more realistic goal for ensuring communicative success, compared to accent reduction or nativelikeness (e.g., Derwing & Munro, 2009; Levis, 2005). However, most of the currently available evidence for the importance of comprehensibility comes from studies that have either treated speakers of various first languages (L1s) as a single group (e.g., Derwing, Munro, & Wiebe, 1998) or focused only on a single L1

(e.g., Munro & Derwing, 1994), making it unclear whether and to what extent linguistic dimensions that feed into comprehensibility and accentedness are specific to the speaker's L1. Therefore, the main objective of the current study was to clarify the relationship between comprehensibility and accentedness, investigating the effect of speakers' L1 on listener perception of L2 speech.

Disentangling Comprehensibility from Accent

For many L2 speakers and their teachers, the ideal ultimate learning goal is often to acquire the linguistic ability of a native speaker (e.g., Derwing, 2003; Levis, 2005; Tokumoto & Shibata, 2011), characterized by native or near-native accent (Derwing & Munro, 2009). However, adult speakers rarely pass for native speakers (Bongaerts, 1999; Bongaerts, Mennen, & Van der Slik, 2000; Moyer, 1999), so accented L2 speech is generally seen as normal and often unavoidable (MacKay, Flege, & Imai, 2006; Major, 2001; Munro, Derwing, & Morton, 2006), even for speakers who begin learning at an early age (Flege, Munro, & MacKay, 1995). Considering the difficulty of acquiring natively like L2 speech, adopting a more realistic learning and teaching goal has been encouraged, with a particular focus on comprehensibility or ease of understanding (Derwing & Munro, 2009; Levis, 2005). Indeed, even a heavy L2 accent does not preclude speakers from being highly comprehensible (Derwing & Munro, 2009; Kang, Rubin, & Pickering, 2010; Munro & Derwing, 1999). And a focus on comprehensibility seems sensible from a practical perspective, given that the interlocutor's goal in most real-world contexts is to get their message across rather than to pass for native speakers. Thus, to make informed decisions about future learning goals and to address these goals through targeted instruction, L2 speakers and their teachers need to know which aspects of language contribute to comprehensible speech and which are tied to the perception of foreign accent.

A focus on comprehensibility (rather than accentedness) is also motivated from a theoretical standpoint. For instance, the Interaction Hypothesis (e.g., Long, 1996) posits that language learning primarily takes place during communication breakdowns in conversations involving L2 speakers. These breakdowns often lead to negotiation for meaning, in which interlocutors make an effort to repair communication through the use of such discourse moves as clarification requests or confirmation checks. According to the Interaction Hypothesis, negotiation for meaning facilitates L2 development by promoting speakers' attention to various linguistic dimensions which may have caused a communication breakdown (for review, see Mackey & Goo, 2007). And because communication breakdowns occur as a result of some linguistic dimensions more so than others (Mackay, Gass, & McDonough, 2000), those dimensions tied to comprehensibility, rather than those that are uniquely linked to accentedness, will be more beneficial in helping learners notice and repair their nontarget production. Thus, to understand which linguistic dimensions of speech are beneficial for L2 development, it is necessary to distinguish the dimensions which contribute to comprehensibility from those that are uniquely tied with accent.

Previous research examining linguistic influences on listener perception of L2 speech has primarily targeted the domains of phonology and fluency. When it comes to understanding, for example, measures such as word stress (Field, 2005), sentence stress (Hahn, 2004), speech rate (Munro & Derwing, 2001), as well as pitch range and pause or syllable length (Kang et al., 2010; Tajima, Port, & Dalby, 1997; Winters & O'Brien, 2013) have all been shown to influence how accurately listeners can extract meaning from an utterance. Although there is little research focusing on domains other than phonology and fluency, poor grammar and inappropriate lexical choice also appear to compromise listener understanding (Fayer & Krasinski, 1987; Munro &

Derwing, 1999). And with respect to accentedness, segmental accuracy (Derwing et al., 1998), pausing and articulation rate (Trofimovich & Baker, 2006), and various suprasegmental measures such as pitch range, stress, and pause length (Kang, 2010) have been linked to listener judgments of L2 accent. In sum, listener judgments of L2 speech, which include comprehensibility and accentedness, are linked to many (often overlapping) linguistic measures spanning the domains of phonology, fluency, grammar, and lexicon.

Moving away from a focus on individual linguistic dimensions, researchers have recently begun to investigate their interactions, targeting the combined contribution of several dimensions to listener judgment. For instance, Trofimovich and Isaacs (2102) recently analyzed the speech of 40 L1 French speakers of English targeting 19 coded linguistic measures (divided into phonology, fluency, lexis/grammar, and discourse-based categories), with the goal of identifying links between these measures and both comprehensibility and accentedness. Comprehensibility was best explained using word stress, type frequency (a measure of lexical richness), and grammar accuracy, while word stress and rhythm best defined accent. A follow-up study, in which the speech of the same speakers was rated for 11 linguistic measures (Saito, Trofimovich, & Isaacs, submitted a), revealed again that comprehensibility was associated with a range of variables, including pronunciation (word stress, speech rate, rhythm), lexis, and grammar, but that accentedness was mainly linked to pronunciation (segmental errors, word stress). A further study targeting 120 L1 Japanese speakers of English similarly showed that comprehensibility was tied to segmental, prosodic, temporal, lexical, and grammatical aspects of L2 speech, while accentedness was related to pronunciation, especially segmental accuracy and word stress (Saito, Trofimovich, & Isaacs, submitted b). Thus, there appears to be consistent evidence that

comprehensibility encompasses a range of linguistic dimensions while accentedness involves mainly pronunciation and fluency factors.

L1 Effects on L2 Comprehensibility and Accentedness

However, one issue that remains to be resolved concerns the extent to which linguistic correlates of comprehensibility and accentedness are specific to speakers' L1 background. On the one hand, the findings reviewed above indicate that comprehensibility and accentedness are distinct constructs, with comprehensibility associated with a broader range of variables. On the other hand, these results show that several variables (e.g., word stress, rhythm) contribute to both comprehensibility and accentedness for speakers from different L1 backgrounds. This raises an interesting question of whether the linguistic variables linked to comprehensibility and accentedness are unique to the particular L1 groups targeted in prior research, or whether at least some of these variables apply equally to speakers from many L1 backgrounds. From a theoretical perspective, there is considerable evidence of L1 effects on L2 development, especially in the realm of pronunciation. Such evidence spans decades of research, starting from early attempts to describe L1 influences on pronunciation as a perceptual "sieve" biasing learners (Trubetzkoy, 1939), to the Contrastive Analysis Hypothesis used to predict and explain speech patterns that may cause particular learning difficulties (Lado, 1957), to more recent conceptualization of L1 effects, including Eckman's (1991) structural conformity hypothesis, Escudero and Boersma's (2004) optimality-theoretic model, Flege's (1995) speech learning model, and Darcy's direct mapping approach (Darcy et al., 2012). Indeed, there appears to be little debate that L1 influence plays a significant role in L2 pronunciation learning (Eckman, 2004), which implies that linguistic correlates of comprehensibility and accentedness might in fact be specific to speakers' linguistic background.

Compared to the vast literature documenting L1 influences on the perception and production of specific aspects of L2 phonology, such as segmental contrasts or voice-onset time (e.g., Davidson, 2011; Major, 2001), there is little research exploring L1 effects on listener-based ratings of comprehensibility and accentedness. The majority of studies considering rater perception of speech have either focused solely on a single L1 group (e.g., Munro & Derwing, 1994; Trofimovich & Isaacs, 2012; Winters & O'Brien, 2013) or conflated multiple L1s into a single group (e.g., Derwing et al., 1998; Kang et al., 2010), and the few studies that have compared different L1 groups have yielded mixed findings. For instance, Anderson-Hsieh, Johnson, and Koehler (1992) found that ratings of speech prosody had a strong positive correlation with L2 speakers' pronunciation scores regardless of their L1 background, whereas segmental and syllable structure errors were dependent on speakers' L1. In terms of other prosodic factors, Baker et al. (2011) reported that word duration and word reduction patterns were negatively associated with accentedness ratings for both Chinese and Korean speakers of English. In contrast, when looking at university-level international teaching assistants, Kang (2010) identified Chinese and Japanese speakers over speakers of other L1s (i.e., Arabic, Russian, Hindi) as having strong accents due to frequent, inappropriate word emphasis. And in a longitudinal study comparing speech ratings of Mandarin and Slavic speakers, Derwing, Munro, and Thomson (2008) found that, while both groups began at equal levels, only the Slavic group showed improvement over time, implying that there could be a possible L1 transfer effect benefiting the Slavic speakers. This limited evidence thus identifies a pressing need to consider how speakers' L1 background affects listener judgments of L2 speech, a point which will be crucial in enabling learners and their teachers to set appropriate learning goals.

The Current Study

In sum, there is growing research interest in identifying linguistic influences on comprehensibility (understanding) and distinguishing such influences from those tied to listener perception of accent (nativeness). The current study contributes to this research agenda by investigating L1 background effects on the relationship between various linguistic dimensions of L2 speech and both comprehensibility and accentedness. The study had two specific objectives: (a) to further clarify which linguistic variables in L2 speakers' speech contribute to the perception of comprehensibility and which to accentedness and (b) to determine if the relative contributions of these linguistic variables differ as a function of speakers' L1 background. To address these objectives, four groups of 15 L2 English speakers from four L1 backgrounds were recorded completing a picture-narrative task, and their speech was subsequently rated for comprehensibility and accentedness and analyzed for 10 linguistic dimensions of their speech. Targeting the same narrative task used in previous research on listener perception of comprehensibility and accent (e.g., Derwing et al., 2008; Trofimovich & Isaacs, 2012) allowed for a direct comparison across studies. And using four typologically different language families (Chinese, Hindi/Urdu, Farsi, Romance) made it possible to document possible L1-specific effects on listener judgments of accent and comprehensibility. Based on the assumption that the needs of L2 speakers from one language background may differ from the needs of those from another, the overall objective of the current study was to provide empirical evidence for language researchers, teachers, and L2 speakers to use in setting appropriate learning goals.

Method

Participants

Speakers. The L2 participants were 60 speakers selected from an unpublished corpus of L2 speech which included audio recordings by 143 speakers from 19 different L1s completing five speaking tasks (Isaacs & Trofimovich, 2011). The speakers, who at the time of the recording were international students in various undergraduate (29) and graduate (31) programs at an English-medium university in Montreal (Canada), were organized in four equal groups ($n = 15$) based on their L1 background (Chinese, Hindi/Urdu, Farsi, Romance). The speakers of Hindi and Urdu were combined into one group because the principal difference between these languages is script-based (King, 1994). The Chinese, Hindi/Urdu, and Farsi groups represented the three largest L1 cohorts in the corpus, with a total of 15, 17, and 32 speakers, respectively. The final group, which comprised the entire set of French and Spanish speakers, served as a comparison group to the French speakers tested by Trofimovich and Isaacs (2012), since Spanish stems from the same language family as French (Posner, 1996) and shares a similar syllable-timed rhythm (Jun, 2005).

In creating the final groups, the speakers in the two larger cohorts (Hindi/Urdu, Farsi) were matched as much as possible to the Chinese and Romance speakers for several background variables (shown in Table 1). The only exception was the 14:1 male-female ratio in the Hindi/Urdu group, which reflected the gender composition of Hindi/Urdu speakers in the larger university community. Despite some minor differences across groups, particularly in self-ratings of speaking and listening ability and self-reported patterns of L2 use (see Table 1), all speakers were deemed to represent a comparable level of L2 oral ability. All speakers had demonstrated at

minimum a speaking score of 17 for TOEFL iBT or 5 for the IELTS proficiency tests, which was considered sufficient for them to pursue academic degrees.

Table 1

Means and Standard Deviations for Participant Background Characteristics

Background variable	Chinese	Hindi/Urdu	Farsi	Romance
Gender (m/f)	6/9	14/1	9/6	9/6
Age	22.5 (2.9)	23.5 (2.0)	25.2 (2.4)	21.4 (3.3)
Years of English study	10.3 (2.9)	14.3 (6.0)	8.5 (4.8)	11.1 (4.3)
Years in Canada	0.7 (.34)	0.4 (.21)	0.4 (.19)	0.6 (.3)
Speaking ability ^a	5.6 (1.2)	7.1 (1.1)	5.7 (1.0)	6.4 (1.1)
Listening ability ^a	5.9 (1.4)	8.0 (0.8)	6.9 (0.9)	7.6 (1.0)
English use at home ^b	17.0 (16.9)	40.0 (26.5)	21.0 (34.1)	29.3 (32.8)
English use at school ^b	72.7 (21.5)	83.3 (20.6)	50.0 (30.5)	79.3 (27.1)

Note. ^aSelf-rating on a 1-9 scale (1 = *extremely poor*, 9 = *extremely fluent*). ^bSelf-rating on a 0-100% scale.

The resulting four L1 groups were considered to provide an appropriate comparison of possible L1 effects on accent versus comprehensibility because these groups represented typologically different languages, belonging to the Sino-Tibetan language family (Chinese) or to the Italic (French, Spanish), Indo-Aryan (Hindi/Urdu), and Iranian (Farsi) sub-branches of the Indo-European language family. The four L1s also crucially differ in their prosody, particularly in terms of rhythm, thus allowing for direct comparisons between the speakers of syllable-timed French tested by Trofimovich and Isaacs (2012) and the speakers of non-Romance syllable-timed Hindi (Dauer, 1983), stress-timed Farsi (Jun, 2005), and tonal Chinese (Jun, 2005).

Raters. The raters included 10 native speakers of English ($M_{\text{age}} = 32.7$ years, $\text{range} = 25\text{-}56$) who were born and raised in English-speaking homes with at least one native English-speaking parent (with seven reporting both parents as native speakers). The raters, who at the time of the study resided in Montreal (a bilingual French-English city), reported speaking English on average 89% of the time ($SD = 8.8$), interacting with native English speakers 73% of the time ($SD = 14.9$), and listening to English media 85% of the time ($SD = 13.5$) daily. The raters had on average 6.6 years of L2 teaching experience ($\text{range} = 1\text{-}23$) and were either enrolled in (9) or recently completed (1) their graduate studies (7 MA, 3 PhD) in applied linguistics or second language education at a local English-medium university. Using a 9-point scale (1 = “not at all familiar”, 9 = “very familiar”), they reported high familiarity with accented English ($M = 8.6$, $\text{range} = 7\text{-}9$), which included L2 speech by speakers of the target languages (i.e., French, Hindi, Chinese). Raters with linguistic and teaching backgrounds were chosen because Saito et al. (submitted a) showed that experienced raters, compared to inexperienced ones, were more consistent in evaluating complex and less intuitive linguistic variables in a similar rating task.

Materials

As part of the original L2 speech corpus, each speaker completed five speaking tasks: a read-aloud task, a picture narrative, an IELTS long-turn speaking task, a TOEFL iBT integrated task, and a Test of Spoken English (TSE) graph-based interpretation task. Of these tasks, the picture narrative was chosen for analysis in the current study because it was the same task used in Trofimovich and Isaacs (2012) and in several similar studies (e.g., Derwing et al., 2008), which allowed for direct comparisons of findings. In the picture narrative task, speakers were presented with an eight-frame colored picture story featuring two strangers bumping into each

other while rounding a corner, then accidentally exchanging their identical suitcases, and finally realizing their mistake upon returning home (Derwing et al., 2008).

All narratives were recorded directly onto a computer using a Plantronics (DSP-300) microphone, stored as digital audio files, and then normalized by matching peak amplitude across files. For each recording, all fillers and false starts at the beginning of the file were removed before it was edited down to the initial 30 s of speech produced, in line with previous research using 20-60 s recordings to evaluate L2 speech (e.g., Derwing et al., 1998; Derwing et al., 2008). All samples were also orthographically transcribed by a trained research assistant and then subsequently verified. The audio recordings and transcripts served as the stimuli for global judgments of accent and comprehensibility as well as for linguistic coding using 10 rated categories spanning the dimensions of phonology, fluency, lexicon, grammar, and discourse.

Speech Rating

All ratings were collected as part of a larger project evaluating speaker performance in three speaking tasks, which included the picture narrative task targeted here. The project involved four individual 2 h sessions (with breaks), all occurring within three weeks of each other, during which the raters evaluated audio recordings or transcripts blocked by task in a counterbalanced order (e.g., Task 1-2-3, 2-3-1, etc.), with audio recordings or transcripts presented to each rater in a unique randomized order. Session 1 was devoted to providing global judgments of accent and comprehensibility based on audio recordings. Session 2 and part of Session 3 were dedicated to rating audio recordings for five phonology- and fluency-based categories. The remainder of Session 3 and Session 4 were spent evaluating orthographic transcripts for five lexical, grammatical, and discourse categories.

All ratings were carried out using a computer-based scale developed by Saito et al. (submitted a), with each measure evaluated on a 1000-point continuous sliding scale and endpoints clearly marked on a horizontal plane (see Appendix for onscreen labels and full training materials). The scales were run through the MATLAB software, and the raters used a free moving slider on a computer screen to assess each category. The rating was recorded as “0” if the slider was placed at the leftmost (negative) end of the continuum, marked with a frowning face. The rating was recorded as “1000” if the slider was set at the rightmost (positive) end of the continuum, marked with a smiley face. On each scale, the slider initially appeared in the middle (corresponding to the rating of 500), and the raters were informed that even a small movement of the slider may represent a fairly large difference in the rating. Apart from brief verbal descriptions for the endpoints of each category and the frowning and smiley faces to indicate the directionality of the scale, no numerical labels or marked intervals were included in the scale.

At the beginning of each session, the raters were seated in front of a personal laptop displaying the rating interface and received training on the relevant rated categories and on the use of the scale. The raters then listened to audio recordings or viewed transcripts and performed four practice judgments by using the appropriate scales, with each practice judgment subsequently discussed between the rater and the researcher to ensure that each measure was understood accurately. The raters were informed that each recording lasted only 30 s, with the possibility that some speakers may have been cut off in the middle of a phrase, and that this should not be taken into consideration when making judgments.

Rated Categories

Accent and comprehensibility. Following Trofimovich and Isaacs (2012), accent was defined as raters’ perception of how different the speaker sounded from a native speaker of

North American English (1 = “heavily accented”, 1000 = “no accent at all”), while comprehensibility was defined as the degree of ease or difficulty in raters’ understanding of L2 speech (1 = “hard to understand”, 1000 = “easy to understand”). Comprehensibility (rather than intelligibility) was chosen as the measure of understanding as it reflects a more typical and practical approach to measuring understanding in a variety of high- and low-stakes assessment contexts (such as oral proficiency scales) and in research settings (Isaacs & Trofimovich, 2012). Because the judgments of accent and comprehensibility reflect quick, intuitive perceptions of L2 speech (Munro, Derwing, & Burgess, 2010; Saito et al., submitted b), the raters were allowed to listen to each recording only once before making their judgment, which ensured that the ratings were comparable across studies.

Phonology and fluency. The raters listened to and evaluated each audio recording for the following five segmental, prosodic, and temporal categories (see Appendix for the full training materials):

1. Segmental errors, defined as errors in the pronunciation of individual consonants and vowels within a word (e.g., *dat* instead of *that*; *pin* instead of *pen*), as well as any segments erroneously deleted from or inserted into words (e.g., *'ouse* instead of *house*; *supray* instead of *spray*).
2. Word stress errors, defined as errors in the placement of primary stress (e.g., *com-pu-TER* instead of *com-PU-ter*, where capital letters designate stress) or the absence of discernible stress, such that all syllables receive equal prominence (e.g., *com-pu-ter* instead of *com-PU-ter*).

3. Intonation, defined as appropriate pitch moves that occur in native speech, such as rising tones in yes/no questions (e.g., Will you be home tomorrow↑) or falling tones at the end of statements (e.g., Yeah, I'll stay at home↓).
4. Rhythm, defined as the difference in stress (emphasis) between content and function (grammatical) words. For instance, in the sentence “They RAN to the STORE”, the words “ran” and “store” are content words and therefore are stressed more than the words “they”, “to”, and “the”, which are grammatical words featuring reduced vowels.
5. Speech rate, defined as the speed with which a speaker produces an utterance.

Unlike accent and comprehensibility ratings, which are quick and intuitive in nature, the judgments of phonology and fluency likely require an in-depth analysis of the speech signal.

Therefore, to ensure the quality of the raters' analysis, they had the option to listen to the same speech sample multiple times until they felt satisfied with their judgment.

Lexicon, grammar, and discourse structure. In order to evaluate each speaker for five categories related to lexicon, grammar, and discourse structure, the raters were presented with written transcripts of each audio recording which had been modified to remove (a) hesitation markers (e.g., um, uh), (b) spelling clues signaling pronunciation-specific errors (e.g., *when* pronounced as *ven* was spelt as intended), and (c) punctuation. This was done to ensure that the raters could make accurate judgments of lexical, grammatical, and discourse-level variables without being overly sensitive to pronunciation and fluency aspects of L2 speech (Patkowski, 1980) and to avoid any transcriber influence (Ochs, 1979). The raters evaluated written transcripts for the following five lexical, grammatical, and discourse categories (see Appendix for full training materials):

6. Lexical appropriateness, defined as the speaker's choice of words to accomplish the task. Poor lexical choices include incorrect, inappropriate, and non-English words (e.g., "A man and a woman bumped into each other on a walkside").
7. Lexical richness, defined as the sophistication of the vocabulary used by the speaker. Simple words with little variety correspond to poor lexical richness (e.g., "The girl arrived home her dog was happy she arrived home", compared to "The girl arrived home to find her dog overjoyed at her return").
8. Grammatical accuracy, defined as the number of grammatical errors made by the speaker. Examples included word order errors (e.g., "What you are doing?"), morphological errors (e.g., "She go to school every day"), and agreement errors (e.g., "I will stay there for five day").
9. Grammatical complexity, defined as the sophistication of the speaker's grammar. Grammatical complexity is low if the speaker uses simple, coordinated structures without embedded clauses or subordination (e.g., "The man wore a black hat and he enjoyed his coffee", compared to "The man that was wearing a black hat was enjoying his coffee").
10. Discourse richness, defined as the richness and sophistication of the utterance content. Discourse richness is low if the entire narrative is simple, unnuanced, bare, and lacks sophisticated ideas or details, but high if the speaker produces several distinct ideas or details so that the statement sounds developed and sophisticated.

As was the case with phonology and fluency judgments, the raters were allowed to spend as much time as needed with each transcript to allow for accurate judgments.

Understanding and use of rated categories. Upon completion of each set of linguistic ratings, the raters used 9-point scales to assess the extent to which they understood the rated

categories (1 = “I did not understand at all”, 9 = “I understand this concept well”) and to which they could comfortably and easily use them (1 = “very difficult”, 9 = “very easy and comfortable”). The raters indicated that they could understand all categories well ($M = 8.3$; $range = 7.8-8.7$) and could use them easily ($M = 7.8$; $range = 7.2-8.3$).

Results

Rater Consistency

The 10 raters were overall consistent in their global judgments of L2 speech, revealing high reliability indexes (Cronbach’s alpha) for accent ($\alpha = .94$) and comprehensibility ($\alpha = .89$). Therefore, mean accent and comprehensibility scores were calculated for each speaker by averaging across all listener ratings. Although the judgments of phonological, lexical, grammatical, and discourse-level categories presumably involve using less intuitive and more complex categories, compared to accent and comprehensibility, the raters were nevertheless fairly consistent, demonstrating reliability indexes that exceeded the benchmark value of .70-.80 (Larson-Hall, 2010) for pronunciation ($\alpha_{segmentals} = .94$; $\alpha_{word\ stress} = .86$; $\alpha_{intonation} = .83$; $\alpha_{rhythm} = .85$), fluency ($\alpha_{speech\ rate} = .90$), vocabulary ($\alpha_{appropriateness} = .80$; $\alpha_{richness} = .87$), grammar ($\alpha_{accuracy} = .81$; $\alpha_{complexity} = .87$), and discourse ($\alpha_{richness} = .89$). One adjustment was made to lexical appropriateness due to a low corrected item-total correlation (.22) specific to one rater by removing this rater’s data ($\alpha_{appropriateness} = .81$). The raters’ scores were therefore considered sufficiently consistent and were averaged across the 10 raters (nine for lexical appropriateness) to derive a single mean score per speaker for each rated category.

Comprehensibility and Accent

The goal of the first analysis was to examine possible group-based differences in global speech ratings. For this, comprehensibility and accentedness ratings were submitted to a two-way

analysis of variance (ANOVA) with group (Chinese, Hindi/Urdu, Farsi, Romance) as a between-subjects factor and perceptual judgment (comprehensibility, accentedness) as a within-subjects factor. The ANOVA yielded a significant main effect of group, $F(3, 56) = 11.29, p < .0001, \eta_p^2 = .38$, a significant main effect of perceptual judgment, $F(1, 56) = 75.5, p < .0001, \eta_p^2 = .57$, and a significant two-way interaction, $F(3, 56) = 3.20, p < .0001, \eta_p^2 = .15$. Tests of interaction effects (Bonferroni adjusted $\alpha = .003$) further showed that the Romance ($p = .002$), Farsi ($p < .0001$), and Hindi-Urdu ($p < .0001$) groups, but not the Chinese group, were rated higher in comprehensibility than in accent, with medium-to-large effect sizes (Cohen's $d = .50$ - 1.22). Tests of interaction effects also revealed that the Chinese group was rated as being more accented than the Romance ($p < .0001$) and the Farsi ($p = .001$) groups, with medium-to-large effect sizes ($d = .56$ - 1.66), and that the Chinese group was rated as being less comprehensible than the other three groups ($p < .0001$), with large effect sizes ($d = 1.68$ - 2.12). In essence, the four groups differed in degree of perceived comprehensibility and accentedness, with the Chinese group overall rated as being less comprehensible and more accented than the remaining groups. Table 2 summarizes global speech ratings for each group.

Table 2

Means and Standard Deviations for Accent and Comprehensibility Ratings by L1 Group (Based on a Continuous 1000-Point Sliding Scale)

L1 group	Accent	Comprehensibility
Chinese	343 (115)	417 (115)
Hindi-Urdu	434 (170)	611 (116)
Farsi	524 (150)	638 (092)
Romance	591 (178)	676 (163)

Ratings of Linguistic Dimensions

The goal of the next analyses was to determine how the global speech ratings of comprehensibility and accentedness related to the 10 rated linguistic categories. First, the linguistic scores for all speakers were submitted to an exploratory Principal Component Analysis (PCA) with Oblimin rotation to determine if the 10 rated linguistic categories showed any underlying patterns based on their clustering. Despite a relatively low sample size ($N = 60$), the Kaiser-Meyer-Olkin value was .85, exceeding the required .60 for sampling adequacy and indicating excellent factorability of the correlation matrix (Hutcheson & Sofroniou, 1999). In addition, a significant Bartlett's test of sphericity, $\chi^2(45) = 692.09, p < .0001$, showed that the correlations between the categories were sufficiently large for PCA. As shown in Table 3, the PCA revealed two factors accounting for 82.3% of total variance. Factor 1, which was labeled "Pronunciation", consisted of the four pronunciation categories, plus speech rate, while Factor 2, labeled "Lexicogrammar", consisted of all vocabulary, grammar, and discourse-level categories and speech rate. Thus, the 10 rated linguistic categories patterned along two separate dimensions (pronunciation and lexicogrammar). Speech rate was common to both dimensions, suggesting that both pronunciation and lexicogrammar are linked to fluency (Segalowitz, 2010).

The resulting pronunciation and lexicogrammar PCA scores, derived through the Anderson-Rubin method of obtaining non-correlated factor scores, were then used as predictor variables in two separate stepwise multiple regression analyses to examine the contribution of pronunciation and lexicogrammar to accent and comprehensibility. While the two regression models accounted for roughly the same total variance (77% for accent, 78% for comprehensibility), the ratio explained by the two factors differed (see Table 4). The variance in

Table 3

Summary of a Two-Factor Solution Based on a Principal Component Analysis of the Ten Rated Linguistic Variables

Factor 1 (Pronunciation)	Word stress errors (.98), Intonation (.94), Segmental errors (.91), Rhythm (.889), Speech rate (.46)
Factor 2 (Lexicogrammar)	Discourse richness (.98), Grammatical complexity (.97), Lexical richness (.97), Grammatical accuracy (.79), Lexical appropriateness (.73), Speech rate (.58)

Note. All eigenvalues > 1

Table 4

Results of Multiple Regression Analyses Using the Factors of Pronunciation and Lexicogrammar as Predictors of Accent and Comprehensibility

Predicted variable	Predictor variables	Adjusted R^2	R^2 change	$F(1, 59)$	p
Accent	Pronunciation	.77	.77	193.39	.0001
Comprehensibility	Pronunciation	.57	.57	79.53	.0001
	Lexicogrammar	.78	.21	104.92	.0001

Note. The variables entered into the regression equation were the two factors obtained in the PCA reported in Table 3.

accent was entirely accounted for by the pronunciation factor (77%), while both the pronunciation (58%) and lexicogrammar (20%) factors contributed to comprehensibility.

L1 Background and Linguistic Dimensions of Comprehensibility and Accentedness

The preceding results have shown that accent is overall related to pronunciation while comprehensibility is linked to both pronunciation and lexicogrammar. The goal of the next analyses was to examine whether this relationship differed as a function of the speakers' L1. First, Pearson correlation coefficients (two-tailed) were computed separately for each group between the two PCA factor scores (pronunciation and lexicogrammar) and comprehensibility and accentedness ratings (Bonferroni adjusted $\alpha = .003$). As shown in Table 5, accentedness was linked to pronunciation for all L1 groups, while the relationship between comprehensibility and the two factor scores differed as a function of group. In particular, comprehensibility was associated with pronunciation for the Chinese group, with lexicogrammar for the Hindi/Urdu group, with both pronunciation and lexicogrammar for the Romance group, and with neither factor for the Farsi group.

Table 5

Pearson Correlations Between the Pronunciation and Lexicogrammar Factors and Accent and Comprehensibility by L1 Group

Factor	Accent				Comprehensibility			
	Chinese	Hindi/ Urdu	Farsi	Romance	Chinese	Hindi/ Urdu	Farsi	Romance
Pronunciation	.75*	.79*	.84*	.93*	.73*	.54	.37	.83*
Lexicogrammar	-.09	.20	-.13	.66	.44	.74*	.49	.78*

Note. * = significant correlation (Bonferroni adjusted $\alpha = .003$).

Table 6

Pearson Correlations Between the 10 Rated Linguistic Categories and Accent and Comprehensibility by L1 Groups

Category	Accent				Comprehensibility			
	Chinese	Hindi/ Urdu	Farsi	Romance	Chinese	Hindi/ Urdu	Farsi	Romance
Segmentals	.91*	.92*	.93*	.96*	.71*	.57	.65	.90*
Word stress	.64	.70	.85*	.90*	.59	.29	.31	.84*
Intonation	.66	.77*	.56	.75*	.64	.52	.13	.69
Rhythm	.51	.52	.63	.95*	.62	.57	.24	.90*
Speech rate	.18	.41	.23	.82*	.67	.60	.33	.88*
Lexical appropriateness	.19	.31	-.06	.60	.18	.77*	.47	.66
Lexical richness	-.15	.27	-.18	.53	.34	.73*	.47	.64
Grammatical accuracy	.08	.18	.13	.71*	.34	.65	.40	.78*
Grammatical complexity	-.22	.18	-.11	.49	.30	.71*	.37	.61
Discourse richness	-.17	.05	-.24	.46	.36	.72*	.36	.61

Note. * = significant correlation (Bonferroni adjusted $\alpha = .003$).

The final analysis further explored which specific linguistic dimensions were associated with comprehensibility and which were linked to accentedness, separately for each L1 group. For this analysis, Pearson correlation coefficients were computed for each L1 group between the 10

rated linguistic categories and comprehensibility and accentedness (Bonferroni adjusted $\alpha = .003$). As summarized in Table 6, accentedness was associated with various rated pronunciation (but not lexicogrammar) categories for all L1 groups, from segmental issues for the Chinese group, segmental issues in combination with intonation or word stress for the Hindi/Urdu and Farsi speakers, to all phonology- and fluency-based categories for the Romance speakers. In contrast, comprehensibility was linked uniquely to segmental issues for the Chinese group, to lexical appropriateness and richness as well as grammatical complexity and discourse richness for the Hindi/Urdu group, and to a combination of segmental, prosodic, and fluency factors, and grammatical accuracy for the Romance group. Confirming an earlier finding, no rated category was found to show a strong association with comprehensibility for the Farsi speakers.

Discussion

Conceptualized as an investigation of L1 influences on the relationship between linguistic dimensions of speech and listener ratings of comprehensibility and accentedness, the current study yielded two main findings. The first finding is that comprehensibility and accent are overlapping yet distinct constructs (e.g., Derwing & Munro, 2009). Consistent with previous research (e.g., Saito et al., submitted a, b; Trofimovich & Isaacs, 2012), accentedness appears to be mainly associated with pronunciation factors, whereas comprehensibility is linked to several pronunciation, fluency, lexis, grammar, and discourse variables. The second main finding is that linguistic dimensions feeding into listener judgments of comprehensibility (but not accentedness) seem to vary as a function of speakers' L1. Accentedness was uniquely linked to pronunciation variables (especially segmental errors) for all L1 groups. In contrast, linguistic correlates of comprehensibility depended on the speakers' L1, such that comprehensibility was associated with pronunciation (segmentals) for the Chinese group, with lexicogrammar (lexical

richness/accuracy, grammar complexity, discourse richness) for the Hindi/Urdu group, with both pronunciation (segmentals, word stress, intonation, rhythm) and lexicogrammar (grammatical accuracy) for the Romance group, and with neither factor for the Farsi group.

Pronunciation Variables and Accentedness

When it comes to accent, the relationship between pronunciation variables and accent ratings is not surprising in light of previous research identifying pronunciation- and fluency-based influences on accent (e.g., Derwing et al., 1998; Kang, 2010; Trofimovich & Baker, 2006). For all four L1 groups, segmental errors represented the strongest influence, consistent with previous studies targeting native speakers of French (Saito et al., submitted a) and Japanese (Saito et al., submitted b). Segmental substitutions in L2 speech may thus be particularly salient to the listener, regardless of the speaker background, making the greatest contribution to perceived accent. However, beyond similarities in segmental influence, each L1 group differed in which of the four remaining pronunciation variables were associated with their accent. For the Romance group, besides segmentals, the two strongest influences were word stress and rhythm, which are often problematic for French and Spanish speakers due to the syllable-timed nature of both languages and the absence of contrastive stress in French (Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Gutiérrez-Díez, 2001). For the Farsi group, word stress had the second strongest link with accent, suggesting that word stress may contribute to accent despite apparent similarities in the stress-timed rhythm between Farsi and English (Dauer, 1983). For the Hindi-Urdu speakers, after segmentals, intonation had the strongest relationship with accent, perhaps due to Hindi/Urdu speakers' use of substantially raised pitch without heavy articulation to indicate emphasis, as compared to the use of pitch in English (Shackle, 2001). And for the Chinese group, segmental errors likely overshadowed other possible influences on accent for the

listener, reflecting a well-documented difficulty Chinese speakers have with English segments (Chang, 2001; Lee, 1976; Rau, Chang, & Tarone, 2009; Zhang & Yin, 2009).

Taken together, these findings suggest that listener perception of accent in L2 speech is the result of a complex interaction between several pronunciation and fluency variables (see Table 6). In judging L2 accent, listeners seem to prioritize segmentals regardless of speaker background but also consider relative weights of other prosody- and fluency-based characteristics of L2 speech, likely based on their perceptual saliency (Goldschneider & DeKeyser, 2001; Kang, 2010). In the current dataset, the Romance and the Chinese speakers represented two possible endpoints in a perceptual continuum of this kind, with listener judgments of accentedness based solely on segmental errors for the Chinese group or on a combination of segmental substitutions and four other pronunciation variables for the Romance group. The Hindi/Urdu and the Farsi groups fell in the middle of this continuum, with only one additional variable (word stress or intonation) beyond segmentals having a strong link to listener judgments of accent. Thus, listeners appear to be able to weigh relative contributions of several pronunciation cues to perceived accent, and listeners' ultimate judgment seems to be determined by the perceptual saliency of these cues, relative to the impact of segmental errors on accent. In future research, it would be important to study how exactly listeners weigh perceptual cues to accent, with the ultimate goal of understanding how listeners deal with multiple cues in making seemingly effortless but highly reliable accent ratings.

Pronunciation Variables and Comprehensibility

With respect to comprehensibility, which has been argued to represent a more realistic L2 learning goal compared to accent reduction (e.g., Derwing & Munro, 2009), the current results revealed clear L1 background effects on listener ratings of comprehensibility. For the Romance

group, both pronunciation and lexicogrammar factors were associated with comprehensibility, consistent with previous research targeting L1 French speakers of English (Saito et al., submitted a; Trofimovich & Isaacs, 2012). However, where the previous studies found associations of lexical, grammatical, and discourse factors with comprehensibility, for this set of speakers, only grammatical accuracy emerged as a factor (Fayer & Krasinski, 1987; Varonis & Gass, 1982). This difference in the impact of lexicogrammar on comprehensibility may be due to the fact that the Romance speakers in this study (all university-level students) represented a narrower range of L2 speaking ability than the French speakers tested earlier (e.g., Saito et al., submitted a), who were L2 users of English from Quebec ranging from complete beginners to balanced bilinguals. Thus, as international students in an English-medium university, the Romance speakers had fewer problems with lexical, grammatical, and discourse-level aspects of L2 speech, compared to at least some L2 users of English from Quebec, which may have reduced the impact of these variables on comprehensibility.

For the Hindi/Urdu group, the relationship between linguistic variables and comprehensibility was restricted to lexicogrammar factors, suggesting that listener-rated ease of understanding was based on these speakers' lexical, grammatical, and discourse-based choices, rather than on the quality of their speech output in terms of pronunciation, fluency, or prosody. This finding underscores a general observation that international students speaking East Indian languages as their L1s, while being noticeably accented, are often proficient speakers of L2 English (Smith, 1992). This finding also highlights the importance of lexicogrammar variables to comprehensibility for speakers who have little difficulty with L2 segments and prosody. In contrast to the Hindi/Urdu speakers, the comprehensibility of the Chinese group was uniquely associated with the pronunciation factor, particularly with segmental accuracy, so that for these

speakers the degree of difficulty in understanding what they are saying was linked solely to accurate production of individual sounds. In essence, listeners prioritized segmental accuracy over lexicogrammar factors in rating these speakers' comprehensibility because segmental errors in L2 English are both problematic for Chinese speakers and also highly salient to listeners (e.g., Chang, 2001; Lee, 1976; Rau, Chang, & Tarone, 2009 Zhang & Yin, 2009).

Finally, for the Farsi group, none of the 10 linguistic variables appeared to bear a strong relationship with these speakers' comprehensibility, suggesting either that the 10 linguistic dimensions failed to capture important influences on comprehensibility for these speakers or that different linguistic dimensions contributed to the comprehensibility of individual Farsi speakers. However, as shown in Table 2, the Farsi speakers were among the two groups rated as being the most comprehensible; they were also the group featuring the least variability in individual speaker's comprehensibility scores. It could also be, then, that for some L2 speakers, such as Farsi speakers in this study, listener-based comprehensibility ratings may be based on a range of variables, with no single variable or cluster of variables bearing particularly strong individual relationship with comprehensibility. The validity of such "distributed" influences on comprehensibility judgments clearly needs to be established in future research.

By way of summary, the findings presented above reinforce the view that speakers' L1 plays an important role in listener-based judgments of L2 comprehensibility, extending previous literature on L1 effects on learning pronunciation (e.g., Davidson, 2011; Eckman, 2004) and other aspects of language (e.g., Lado, 1957 Weinrich, 1953). The current findings also illustrate several distinct patterns of linguistic influences on L2 comprehensibility: from individual pronunciation or lexicogrammar effects for the Chinese and Hindi/Urdu speakers, to combined pronunciation and lexicogrammar effects for the Romance speakers, to weak or potentially

“distributed” effects of many variables for the Farsi speakers. Considering that the two language groups where comprehensibility was associated with lexicogrammar variables involved syllable-timed languages (Romance, Hindi/Urdu), compared to the stress-timed Farsi and the tonal Chinese, it could be that the impact of lexicogrammar on L2 comprehensibility may be closely tied to prosody (and especially word stress and rhythm). Put differently, prosody and lexicogrammar factors may be closely tied together in their impact on comprehensibility, especially for speakers of syllable-timed languages, such as French, Spanish, and Hindi/Urdu (as in this study and in Trofimovich & Isaacs, 2012), or speakers of moraic Japanese (as in Saito et al., submitted b), who are learning L2s with stress-timing (vowel reduction) or contrastive stress. In this case, the use of stress and rhythm to produce lexical phrases and also the use of pitch to mark tonal groups, phonological phrases, or syntactic phrasal boundaries might serve as important determinants of L2 comprehensibility – through their association with lexical and grammatical appropriateness and sophistication – for speakers from various L1 backgrounds. Clearly, this prosody-lexicogrammar link needs to be studied further in future research.

Limitations and Future Research

Although the importance of several linguistic variables to listener perception of L2 comprehensibility appears strong, as the current findings suggest, potential effects of other factors cannot be ignored. For instance, the Farsi speakers in this study showed no association between comprehensibility and any of the 10 linguistic dimensions analyzed, leaving the possibility that, at least for some L2 speakers, variables outside the scope of this study may impact communicative success. It may be that the targeted linguistic variables were too broad to capture any actual influence. Compared to the 10 rated linguistic dimensions targeted here, 19 different variables were examined by Trofimovich and Isaacs (2012), providing a more finite

view of specific measures of phonology (e.g., pitch contour and range, syllable structure), fluency (e.g., rate of filled and unfilled pausing, pause errors), lexis/grammar (e.g., self-corrections, token frequency, type frequency), and discourse (e.g., story cohesion, breadth, and depth). Therefore, for the Farsi speakers in particular and for L2 speakers at higher proficiency levels in general, a more nuanced set of measures might be needed to identify linguistic correlates of comprehensibility.

Another potential factor influencing the current findings may be listener bias, as listeners tend to rapidly adapt to foreign-accented speech upon exposure to it (e.g., Bradlow & Bent, 2008; Clarke, 2004; Sidaras, Alexander, & Nygaard, 2009). Because all raters in this study were either current or recently graduated students residing in Montreal, they likely had a high level of familiarity with Farsi-accented English due to a large number of Farsi-speaking students in the targeted participant pool (34/143 or 22%) and also with French-accented English since at least 10 speakers from the Romance group also spoke French. Thus, the finding that the Farsi and the Romance speakers were among the most comprehensible and least accented in this study (Table 2) may be at least in part due to listener experience effects on perceptual judgments of speech. In fact, it could be argued that the similarity between Farsi and English in stress-timing, compared to differences between syllable-timed French and Spanish and stress-timed English, may have made it more difficult for the listeners to isolate any specific variable as a major contributor to the comprehensibility of Farsi speakers. Therefore, potential interactions between speakers' L1 background and listeners' experience effects emerge as an interesting target of future research into listener judgment of L2 speech.

Another limitation of this research concerns the focus on L2 speakers who had reached a certain benchmark of L2 proficiency, represented here by minimum TOEFL or IELTS scores

required for their acceptance to an English-medium university. Saito et al. (submitted b) showed that linguistic influences on comprehensibility and accentedness can vary as a function of speaker L2 proficiency level. For example, optimal rate of speech, appropriate and rich vocabulary use, and good prosody may be important for beginner-to-intermediate comprehensibility levels whereas segmental accuracy, good prosody, and correct grammar may matter for intermediate-to-advanced comprehensibility. Therefore, future research with L2 speakers who come from various L1 backgrounds and differ in ability levels is necessary to confirm whether the current findings remain consistent or vary across proficiency levels.

Yet another limitation of this research pertains to the specific nature of the task used (picture narrative). Previous research has shown that L2 speakers tend to score higher in fluency ratings for monologue- and dialogue-based tasks than they do for picture narratives (Derwing, Rossiter, Munro, & Thomson, 2004), and that narrative tasks tend to generate more accuracy with less complexity, compared to tasks that require some form of decision, such as giving advice to people with personal problems (Skehan & Foster, 1997). For this reason, future studies targeting monologue, interview, and two-way interaction tasks may reveal potential task effects on listener judgment of comprehensibility and accentedness. Last but not least, future research must also consider the type of rater performing evaluations of L2 speech. Given that the majority of the world's interactions in English take place between non-native interlocutors (Tucker, 1998), it is important to understand whether the linguistic variables influencing native listener perception of comprehensibility and accentedness will be similar to those affecting non-native listeners.

Conclusion

The findings of this study point to two broad conclusions. The first is that, when judging L2 comprehensibility, native-speaking listeners consider not only pronunciation- and fluency-related aspects of L2 speech, but also grammatical, lexical, and discourse-based variables. All these variables likely determine the time and effort needed for listeners to extract meaning from L2 speech (Munro & Derwing, 1995). This is in contrast to listener judgment of L2 accent, where ratings appear to be invariably fast, effortless, and intuitive (Munro et al., 2010) and primarily linked to segmental aspects of L2 speech (Saito et al., submitted a, b). And the second conclusion is that linguistic influences on comprehensibility vary as a function of speakers' L1, with some speaker groups judged as being comprehensible solely on the basis of their pronunciation or lexicogrammar factor and some groups judged on the basis of several linguistic factors in combination.

Although broad implications of these findings are premature, there are nevertheless several promising suggestions for language teaching. For example, instructors teaching L2 learners who share the same L1 background may benefit from an understanding of the specific linguistic variables that impact their learners' comprehensibility and accentedness. And while instructors teaching learners from multiple L1 groups cannot take full advantage of L1-specific knowledge, they can still promote learners' communicative success by expanding their teaching targets beyond segmentals (e.g., Munro & Derwing, 2006), to include syllable structure (e.g., Couper, 2006), word stress (e.g., Field, 2005), or fluency phenomena (e.g., Munro & Derwing, 2001). This change in focus, along with instruction on how fluency affects comprehensibility (Derwing et al., 2004), how grammatical errors in L2 speech are perceived by listeners (Derwing, Rossiter, & Ehrensberger-Dow, 2002), and how lexical knowledge is linked to listener

understanding (Saito et al., submitted b) should all lead to communicative improvement. In essence, targeting L2 comprehensibility as a teaching and learning goal requires a comprehensive approach sensitive to the variety of L1s that exist within a language classroom.

Chapter 3

The results of the above study yielded two main findings. The first adds support to previous research indicating that comprehensibility and accent are overlapping yet distinct constructs (e.g., Derwing & Munro, 2009). The second suggests that the linguistic dimensions that influence how L2 comprehensibility is perceived vary as a function of speakers' L1. These findings would seem to offer promising suggestions for language teaching, although more information is still needed before making any broad statements on overall implications. This chapter takes a more in-depth look at two potentially rewarding lines of research stemming from the results of this study, the first being whether task affects how L2 speech is perceived, and the second looking at how results may differ for nonnative listeners.

A key limitation to this study is the singular nature of the task analyzed. All participants completed a picture narrative, a common task in L2 speech research (e.g., Derwing et al., 2004, 2008; Saito et al., submitted a; Trofimovich & Isaacs, 2012). In a study comparing fluency ratings across tasks, Derwing et al. (2004) found that L2 speakers tended to score higher in their fluency ratings on monologue- and dialogue- based tasks than on the picture narrative task featured in this study. A key reason for this difference in fluency may be the vocabulary limitations of a picture narrative task. In order to accurately relay the story, participants must use the items in the pictures, and thus it is likely that their perceived disfluency is linked to a lexical search for vocabulary they are less familiar with (Hilton, 2008). Such restrictions are not as prevalent in monologue- and dialogue-based tasks. In another study on fluency, Ejzenberg (2000) showed that L2 speakers were perceived as being more fluent when completing dialogue- than monologue-based tasks, with variation within task due to the amount of direction they

received (an uncued dialogue was perceived as being more fluent than a cued one, with a cued monologue being more fluent than an uncued one).

Though an extremely limited sample, these two studies would seem to indicate a hierarchy of task effects on fluency, with dialogue tasks perceived as the most fluent, followed by monologue tasks, and with picture narratives as the least fluent. Although these results are specific to fluency, task effects on comprehensibility judgment cannot be ignored, as the findings from the current study indicate the importance of fluency in L2 speech. Following the PCA analysis, speech rate was the only linguistic measure that was found to contribute to both the pronunciation and lexicogrammar factors. Furthermore, an association between speech rate and comprehensibility was found for the Romance speakers of this study, as well as in previous studies looking at French speakers (Saito et al., submitted a) and Japanese speakers (Saito et al., submitted b). Considering this existing association (at least for some L1 speakers) in a task that appears to create fluency difficulties for L2 speakers, an investigation into whether such an association exists in tasks that tend to produce greater fluency could lead to different results.

A similar argument can be made for the role of prosody for these same learners (Romance, French, Japanese). In a training study, Derwing et al. (1998) found that while two groups receiving either segmental or prosody-based pronunciation instruction showed improvement in a sentence read-aloud task, only the prosody group showed improvement in comprehensibility and fluency in a picture narrative task (the same as the one used in this study), indicating that prosody-based variables may have a much greater influence in tasks that require speakers to formulate their own utterances. Understanding how the strength of influence that prosodic factors and fluency have on comprehensibility varies as a result of task for L2 speakers

would allow language researchers and teachers to estimate how much weight to place on these factors within pronunciation instruction.

Beyond prosody and fluency, several other linguistic variables measured in this study have received attention to determine if a task effect exists. While focusing on effects of task planning, Skehan and Foster (1997) also found that narrative-based tasks generated more accuracy and less complexity in the language learners produced, while tasks that required some form of decision (i.e., giving advice to people with personal problems) generated more complexity but less accuracy. As grammatical complexity and accuracy were both strong influences on the perception of comprehensibility for Romance and Hindi/Urdu speakers in this study, determining whether the strength of this variable fluctuates as a result of task seems a necessary next step. Besides investigating whether the relative weight of these associations (fluency, prosody, grammar, lexis) with comprehensibility varies as a result of task, further research may also reveal associations between these linguistic measures and comprehensibility that did not exist for L1 groups (i.e., Chinese, Farsi) in the picture narrative task.

A further consideration is the relevance of the task used for analysis in this study. Though the task served to draw comparisons with previous research, as the target population of this study was university students, a picture-narrative task may not be the most revealing. Tasks that are centered around communicative skills that students need to successfully complete for their university degree would provide more depth to this line of research, specifically in formulating approaches towards pronunciation instruction. Such tasks would likely target learners' ability to express their opinion and experience, spontaneously interpret data within a group setting, and integrate reading and listening skills with speaking. Variations of such tasks are available through several university and occupational proficiency exams, including the International

English Language Testing System (IELTS), the Test of Spoken English (ETS), and the Test of English as a Foreign Language (TOEFL). To make any claims about the importance of any of the linguistic variables of interest for university-level students, the information gathered should give an accurate, rounded view on the communicative needs of these students, and thus a focus on these types of tasks would be the next logical line of inquiry.

Beyond task, another limitation is that all raters in this study were native speakers (NS) of English. While this would be ideal for learners studying and working in contexts where they interact primarily with NS, this does not reflect a growing trend across the globe. With multilingualism being common and even prevalent in the majority of communities at the societal level, implying the existence of two or more languages within a speech community (Gorter et al., no date), it is clear that bilingualism and multilingualism constitute an everyday experience for the vast majority of world citizens (Tucker, 1998). It would seem, then, that if the target of pronunciation instruction is communicative success, then we have to expand our views on the perception of comprehensibility and accent beyond NSs to include non-native speakers (NNS) as well.

As seen in the results, the language groups rated as being most comprehensible and least accented were Romance and Farsi, the two groups best represented by the general population where the 10 raters of the study resided. This is not a surprising result as previous literature has indicated that accent familiarity can, to different degrees, affect the perception of L2 accent (e.g., Bradlow & Bent, 2008; Winke & Gass, 2013; Winke, Gass, & Myford, 2013). Many of these studies have viewed this familiarity from the perspective of NS experience. From the perspective of a NNS, the findings are less clear. Major, Fitzmaurice, Bunta, and Balasubramanian (2002) reported that Japanese and Chinese listeners found Spanish-accented English more

comprehensible than accented speakers from their own language background, whereas the Spanish listeners found speakers from their L1 easier to comprehend. In fact, for the Japanese and Chinese listeners Spanish-accented English was almost as comprehensible as that of NSs. This would seem to indicate that Spanish speakers of English are likely to be perceived as more comprehensible than other L2 speakers by listeners of various L1 backgrounds. Munro et al. (2006) found a slight advantage for Japanese over Mandarin and Cantonese listeners hearing Japanese-accented English, but no such advantage existed for Cantonese listeners hearing Cantonese-accented English. In terms of overall ratings, the Japanese listeners rated speakers from their own L1 as easier to understand than Cantonese-accented speech, but not Polish or Spanish-accented speech. The Cantonese group rated Cantonese-accented speech as easier to understand than Japanese, Polish, and Spanish. The final group, Mandarin, found Japanese- and Cantonese-accented speech easier to understand than Spanish. These findings seemingly contradict those of Major et al., who found Spanish-accented English as the easiest to comprehend.

Clearly, these conflicting results, as well as lack of depth in research in this area of speech perception, indicate a necessity to investigate further how the L1 background of NNS speakers affects their perception of L2 speech. When Jenkins (2002) proposed the Lingua Franca Core, the intent was to develop an approach towards pronunciation instruction that would allow for mutual understanding between NNS of English. However, as evident by the results of this study, the linguistic variables that feed into NS perception of L2 speech vary as a result of L1. If variation exists in the perception of L2 speech between NNS listeners from different L1 backgrounds, then it seems likely that variation will also exist amongst the linguistic measures that influence this perception.

Although a focus on comprehensibility as an ultimate learning goal has been encouraged within L2 pronunciation literature (e.g., Derwing & Munro, 2009; Levis, 2005), as reviewed above, many questions remain in determining how best to accomplish this goal. Beyond identifying the linguistic measures that influence the perception of speech (e.g., Saito et al., submitted a, b; Trofimovich & Isaacs, 2012) and demonstrating how these measures vary as a function of learners' L1 (as investigated within this study), more attention needs to be devoted to both the nature of the speaking task as well as the audience of listeners it is being delivered for. To simply state that pronunciation instructors should place a focus on communicative success does not give justice to the intricacies of how L2 speech is perceived. By emphasizing L1 background effects on L2 speech perception, this study attempted to further our knowledge of the interplay between these intricacies, and the results, while promising, await support from future research exploring the role of task and rater background before more concrete steps can be taken in creating a classroom approach to pronunciation instruction.

References

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529-555.
- Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. (2011). Word durations in non-native English. *Journal of Phonetics, 39*, 1-17.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late learners of Dutch as a second language. In D.P. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-59). Mahwah, NJ: Lawrence Erlbaum.
- Bongaerts, T., Mennen, S., & Van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica, 54*, 298-308.
- Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707-729.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching and pronunciation: A course book and reference guide* (2nd ed.). New York, US: Cambridge University Press.
- Chang, J. (2001). Chinese speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd Ed., pp. 310-324). Cambridge, UK: Cambridge University Press.
- Clarke, C. M. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*, 3647-3658.

- Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*, 46-66.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.
- Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. *Second Language Research, 28*, 5-40.
- Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass, 5*, 126-139.
- Derwing, T. M. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review, 59*, 476-490.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 20*, 1-16.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*, 592-603.
- Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29*, 359-380.
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393-410.

- Derwing, T. M., Rossiter, M. J., & Ehrensberger-Dow, M. (2002). "They speaked and wrote real good": Judgements of non-native and native grammar. *Language Awareness, 11*, 84-99.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*, 655-679.
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustic Society of America, 110*, 1606-1618.
- Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*, 23-41.
- Eckman, F. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition, 26*, 513-549.
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 287-313). Ann Arbor, US: The University of Michigan Press.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*, 551-585.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313-326.
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399-423.

- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, US: York Press, 1995.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*, 1-26.
- Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*, 1-22.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning, 51*, 1-50.
- Gorter, D., Cenoz, J., Nunes, P., Riganti, P., Onofri, L., Puzzo, B., & Sachdeva, R. (no date). Benefits of linguistic diversity and multilingualism. Cultural diversity as an asset for human welfare and development. *Sustainable Development in a Diverse World. Eurodivision*, 1-45. Accessed at: http://www.susdiv.org/uploadfiles/RT1.2_PP_Durk.pdf.
- Gutiérrez-Díez, F. (2001). The acquisition of English syllable timing by native Spanish learners of English: An empirical study. *International Journal of English Studies, 1*, 93-113.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201-223.

- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal*, 36, 153-166.
- Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist*. London, UK: SAGE Publications Ltd.
- Isaacs, T., & Trofimovich, P. (2011). *International students at Canadian universities: Validating a pedagogically-oriented pronunciation scale*. Unpublished corpus of second language speech.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34, 475-505.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23, 83-103.
- Jun, S-A. (2005). Prosodic typology. In Sun-Ah Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford, UK: Oxford University Press.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301-315.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94, 554-566.
- King, C. (1994). *One language, two scripts: The Hindi movement in nineteenth century North India*. Delhi, India: Oxford University Press.

- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, US: The University of Michigan Press.
- Lee, M-S. (1976). Pronunciation problems are different and similar among Cantonese and Mandarin speakers. *TESL Reporter*, 9(4), 3-6.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39, 369-377.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition: Second language acquisition* (pp. 413-468). New York, US: Academic Press.
- MacKay, I. R. A., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, 27, 157-183.
- MacKay, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471-497.
- MacKay, A. & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. MacKay (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford, UK: Oxford University Press.
- Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum.

- Major, R. C., Fitzmaurice, S. F., Bunta, F., Balasubramanian, C. (2001). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36, 173-190.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21, 81-108.
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20, 38-51.
- Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11, 253-266.
- Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285-310.
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451-468.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520-531.
- Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, 44, 316-327.
- Munro, M. J., Derwing, T. M., & Burgess, C. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52, 626-637.

- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Ochs, E. (1979). Transcription as theory. In E. Orchs and B. B. Schieffelin (eds.), *Developmental pragmatics* (pp. 43-72). New York, US: Academic Press, Inc.
- Ortega, L. (2009). *Understanding second language acquisition*. London, UK: Hodder Education.
- Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449-472.
- Posner, R. (1996). *The romance languages*. Cambridge, UK: Cambridge University Press.
- Rau, D. V., Chang, H-H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning*, 59, 581-621.
- Saito, K., Trofimovich, P., & Isaacs, T. (submitted a). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*.
- Saito, K., Trofimovich, P., & Isaacs, T. (submitted b). Developing a process-oriented model for linguistic influences on comprehensibility and accentedness in second language speech production. *Applied Psycholinguistics*.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, US: Routledge.
- Shackle, C. (2001). Speakers of South Asian languages. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd Ed., pp. 310-324). Cambridge, UK: Cambridge University Press.

- Sidas, S., Alexander, J., & Nygaard, L. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*, 3306-3316.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*, 185-211.
- Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 75-90). Chicago, US: University of Illinois Press.
- Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*, 1-24.
- Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes*, *30*, 392-408.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1-30.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905-916.
- Trubetzkoy, N. S. (1939). *Grundzüge der phonologie*. Los Angeles, US: University of California Press.
- Tucker, G. R. (1998). A global perspective on multilingualism and multilingual education. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 3-15). Clevedon, UK: Multilingual Matters.
- Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, *4*, 114-136.

- Weinreich, U. (1953). *Languages in contact: findings and problems*. New York, US: Linguistic Circle.
- Winke, P., & Gass, S. (2013). The influence of L2 experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, *47*, 762-789.
- Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*, 231-252.
- Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication*, *55*, 486-507.
- Zhang, F., & Yin, P. (2009). A study of pronunciation problems of English learners in China. *Asian Social Science*, *5*, 141-146.

Appendix

Training materials and onscreen labels for all perceptual judgments

A. Global judgment	
Accentedness	<p>This refers to how much a speaker's speech is influenced by his/her native language and/or is coloured by other non-native features.</p> <p style="text-align: center;">1 ○-----○ 1000</p> <p style="text-align: center;">“heavily accented” “no accent at all”</p>
Comprehensibility	<p>This term refers to how much effort it takes to understand what someone is saying. If you can understand with ease, then a speaker is highly comprehensible. However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker has low comprehensibility.</p> <p style="text-align: center;">1 ○-----○ 1000</p> <p style="text-align: center;">“hard to understand” “easy to understand”</p>
B. Phonology and fluency judgment	
Vowel and consonant errors	<p>This measure applies to individual sounds and refers to errors in the pronunciation of individual sounds within a word. These errors may affect both consonants and vowels:</p> <ul style="list-style-type: none"> ○ Speaker says “that” but you hear <i>dat</i> ○ Speaker says “pen” but you hear <i>pin</i> <p>Such errors also include the removal and additions of sounds:</p> <ul style="list-style-type: none"> ○ Speaker says “house” but you hear <i>ouse</i> ○ Speaker says “spray” but you hear <i>supray</i> <p style="text-align: center;">1 ○-----○ 1000</p> <p style="text-align: center;">“frequent” “infrequent or absent”</p>
Word stress errors	<p>This measure applies to individual words that are longer than one syllable and refers to errors in the placement of stress in words with more than one syllable. These errors include misplaced stress:</p> <ul style="list-style-type: none"> ○ <i>comPUter</i> is pronounced as <i>compuTER</i> ○ <i>FUture</i> is pronounced as <i>fuTURE</i> <p>These errors also include absent stress, so all syllables sound the same:</p> <ul style="list-style-type: none"> ○ <i>comPUter</i> is pronounced as <i>computer</i> ○ <i>FUture</i> is pronounced as <i>future</i> <p style="text-align: center;">1 ○-----○ 1000</p> <p style="text-align: center;">“frequent” “infrequent or absent”</p>
Intonation	<p>This measure applies to utterances longer than a single word and can be described as the melody of speech. It refers to natural movements of pitch as we produce utterances.</p> <ul style="list-style-type: none"> ○ Pitch goes up in <i>Will you be home tomorrow</i>↑? ○ Pitch goes down in <i>Yeah, I'll stay at home</i>↓ ○ Pitch goes down and up in <i>I'll stay at home</i>↓↑... <i>but only until 3.</i> <p>Intonation should come across as natural and unforced.</p> <p style="text-align: center;">1 ○-----○ 1000</p> <p style="text-align: center;">“unnatural” “natural”</p>

Speech rate	<p>This measure applies to utterances and describes how slowly or quickly someone speaks.</p> <ul style="list-style-type: none"> ○ Speaker can speak slowly with many pauses and hesitations. ○ Speaker can speak very fast. ○ Speakers can speak at a natural rate and can be comfortable to listen to. <p>1 ○-----○ 1000 “too slow or too fast” “optimal”</p>
Rhythm	<p>This measure applies to utterances and refers to differences in stress (emphasis) between content and function (grammatical) words.</p> <ul style="list-style-type: none"> ○ In <i>they RAN to the STORE</i>, the words <i>ran</i> and <i>store</i> are all content words and therefore are stressed more than the words <i>they</i>, <i>to</i> and <i>the</i>, which are grammatical words. <p>Rhythm should sound and feel natural in speech.</p> <p>1 ○-----○ 1000 “unnatural” “natural”</p>
C. Lexical, grammatical, discourse-level judgment	
Lexical appropriateness	<p>This measure applies to individual words and refers to a speaker’s choice of words to accomplish a speaking task. Poor lexical choices include incorrect, inappropriate, and non-English words.</p> <ul style="list-style-type: none"> ○ <i>I drank coffee with my friends in a fancy French <u>cafeteria</u>.</i> ○ <i>A man and a woman bumped into each other on a <u>walkside</u>.</i> <p>1 ○-----○ 1000 “many inappropriate words used” “consistently uses appropriate vocabulary”</p>
Lexical richness	<p>This measure applies to individual words and refers to the sophistication of the vocabulary used by a speaker to discuss a particular topic. Lexical richness is poor if a speaker uses very simple words with little variety.</p> <ul style="list-style-type: none"> ○ More rich utterance: <i>The girl arrived home to find her dog overjoyed at her return she quickly realized that he was more likely excited for the cookie he was about to receive.</i> ○ Less rich utterance: <i>The girl arrived home her dog was happy she arrived home and the dog was happy too because he could eat a cookie.</i> <p>1 ○-----○ 1000 “few, simple words used” “varied vocabulary”</p>
Grammatical accuracy	<p>This measure applies to both individual words and utterances longer than a single word and refers to the number of grammar errors made by the speaker. These may include:</p> <ul style="list-style-type: none"> ○ Errors of word order: <i>What you are doing?</i> ○ Errors in grammar endings: <i>She go to school every day.</i> ○ Agreement errors: <i>I will stay there for five day.</i> <p>1 ○-----○ 1000 “poor grammar accuracy” “excellent grammar accuracy”</p>

Grammatical complexity	<p>This measure applies to utterances that are longer than a single word and describes the complexity and sophistication of a speaker's grammar. Grammar is sophisticated if a speaker uses complex and elaborate structures and embeds shorter utterances within longer utterances.</p> <ul style="list-style-type: none"> ○ More complex utterance: <i>The man that was wearing a black hat was greatly enjoying his coffee.</i> ○ Less complex utterance: <i>The man wore a black hat... and he enjoyed his coffee.</i>
1 ○-----○	1000
“simple grammar”	“elaborate grammar”

Discourse richness	<p>This measure applies to the entire narrative and describes how rich and sophisticated a speaker's narrative is.</p> <ul style="list-style-type: none"> ○ Discourse richness is low if the narrative is simple, unnuanced, bare, and lacks sophisticated ideas or details. ○ Discourse richness is high if a speaker produces several distinct ideas or details in his or her narrative, so that the story sounds developed and sophisticated.
1 ○-----○	1000
“simple structure, few details”	“detailed and sophisticated”
