TRENDS IN ACCENTEDNESS AND COMPREHENSIBILITY RESEARCH, WITH RESPECT OF L2 SPEECH RATINGS: A LITERATURE REVIEW

LIVIA POLJAK

Simon Fraser University

Abstract

Accentedness and comprehensibility research has greatly evolved since its infancy in in the 1980s. As language learner trends have shifted from native-like mastery to being comprehensible to listeners, a broader range of factors, such as speaker rates, and lexical/grammatical measures have since been evaluated for research and pedagogical purposes. This paper attempts the break down the major methodological procedures used in accentedness and comprehensibility research, chiefly, by examining the types of speakers and listeners that are most commonly present in studies, the types of stimuli and rating methods, the various features of accentedness and comprehensibility and listener attitudes, to name a few. Finally, this review briefly highlights what new components should be added in order to further our understanding in this field, both for academic interest and practical pedagogical purposes.

Keywords: standards-based grading, assessment, motivation, affect, mathematics, classroom

Introduction

The shift from preferred monolingualism to encouraged multilingualism, both in schools and the global community, has been, in the past few decades, ongoing if not slow. The early 1980s had seen the rise of bilingual education systems erected in schools (of these, most notable are the French immersion school in Canada and the Spanish bilingual schools found throughout the United States). Yet these systems, as well as many other second language (L2) classrooms, still follow in the footsteps of monolingualism, in which students are expected to learn the language, ideally, like a monolingual speaker of the target language (Kramsch, 2009).

These idealizations often ignore the context in which individuals either use their L2, or have acquired it. For example, a speaker's multilingual heritage or their unwillingness to use their first language (L1) do not come into focus in the language classroom as it does on the academic researcher's laptop. This is because the language classroom is, above everything, an evaluation based institution, like any other subject in a school. The target language is relayed by instructors to students who are then evaluated (often in the form of grammar tests or written essays) in order to assess if the target material has been learned. Often pre-existing curriculums set by provincial or federally mandated bodies (such as ministries of education) dictate how a language can be evaluated, and very often, native or nativelikeness is the point of reference. Whether it be students in a bilingual school program, or adult English as a Second Language (ESL) learnings taking classes for proficiency exams such as IELTS or TOEFL, written linguistic competence has always been the major area of interest for Second Language Acquisition (SLA) researchers, which is why evaluations of written grammatical lexical "correctness" have much clearer evaluation techniques, and why methods of evaluating learners' accents or 'accentedness', both in the classroom, and in the research field, are not as clearly defined (see, for example, Genesee's, 1978, explanation on the lack of research on accents in the French immersion classroom).

Indeed accentedness, nativelike pronunciation and even oral comprehensibility have been ignored in the classroom when it came to grading students. However, it should also be noted that from the early twentieth century until about the 1980s, foreign or non-native accent eradication was viewed by researchers and second language teachers as a key objective for L2 speakers of a language, believing that improvement or even the elimination of non-native accents was imperative to comprehensibility (Lippi-Green, 1997; Munro, 2003). More recently though, and especially in the mid 1990s through the work of Munro and Derwing (see: 1995, 1999), multiple factors outside of pronunciation have become targets (such as segmental measures, optimal speaking rate and lexicogrammar) for a more multi-faceted approach to L2 speech. On the whole, as sounding perfectly nativelike was eclipsed by sounding more comprehensible to interlocuters, a more joint approach to accentedness, comprehensibility, and at times other factors such as intelligibility and fluency, have become the basis for L2 speech research. However, while many studies have come to similar conclusions about non-native speech, there is still little consensus in the applied linguistics community about optimal rating scales for measuring speakers or speech features, participant types selected for analysis, or even rater types

used for evaluating speakers. In general, the studies also tended to focus on English as a target L2, which does seem to overestimate the global relevance of English as a world language. Though studies on accentedness and comprehensibility are still evolving, the purpose of this analysis is to observe the direction of this research community with respect to modern trends in language learning.

In this paper, I aim to review, categorize and summarize many of the studies on accentedness and comprehensibility that have been written to date. I will discuss the methodological procedures used by researchers, including the selection of speaker and listener types employed for research, as well as rating tools and stimuli. The variety (or in this case, the lack-there-of) in L2 types will also be discussed. My primary goal is to assess the direction we, as language education researchers, have been headed in the field of SLA research. In my concluding thoughts, I will present what this form of research may be lacking, and what potential changes would need to occur if research on comprehensibility is to advance, in the future.

Some Methodologies Used in Recent Research on Accentedness and Comprehensibility

Speakers

Though L2 backgrounds have differed considerably in accent and comprehensibility studies, L2 English is by far the most common language that was analysed, with, perhaps the exception of the 2014 O'Brien study on L2 German (for more information on this, and all other studies listed here, please see the summaries in Appendix 1). The speakers' 20 different L1 used in the 36 research studies analyzed for this paper are summarized in Table 1.

Speaker L1	**Research Authors**
Mandarin	Anderson-Hsieh & Koehler (1988), Ballard (2013), Crowther, Trofimovich, Saito & Isaacs (2015), Crowther, Trofimovich, Isaacs, & Saito (2015), Isaacs & Thomson (2013), Kang (2010), Kang, Rubin, & Pickering, (2010), Kennedy & Trofimovich (2008), Kennedy, Foote, & dos Santos Buss (2015), Kim (2008), Munro & Derwing, (1995a), Munro & Derwing (1995b), Munro & Derwing (2001), Munro, Derwing, & Morton (2006), Trofimovich, Isaacs, Kennedy, & Saito (2016)
Korean	Kang (2010), Kim (2008), Kang et al., (2010), Sereno, Lammers, Jongman (2016), Trofimovich & Baker (2006)
French	Crowther, Trofimovich, Isaacs, et al. (2015), Isaacs & Trofimovich (2012), Isaacs & Thomson (2013), Kim (2008), Trofimovich & Isaacs (2008), Saito, Trofimovich, & Isaacs (2016), Saito, Webb, Trofimovich, & Isaacs (2016a), Saito, Webb, Trofimovich, & Isaacs (2016b), Trofimovich, et al., (2016)
Russian	Munro & Derwing (2001), Kang (2010),
Hindi/Urdu	Crowther, Trofimovich, Saito, & Isaacs (2015), Crowther, Trofimovich, Isaacs, et al. (2015), Kang (2010), Trofimovich, et al. (2016),
Arabic	Ballard (2013), Kang (2010), Kang, Rubin, Pickering (2010), Kim (2008), Munro & Derwing (2001)
Serbo-Croatian	Isaacs & Thomson (2013), Kang (2010), Munro & Derwing (2001)
Spanish or Spanish-Basque	Burda, (2000), Crowther, Trofimovich, Isaacs, et al. (2015), Derwing & Munro (1997), Isaacs, et al. (2014), Kim (2008), Kang, et al. (2010),

	Munro & Derwing (2001), del Puerto, Lacumberri, & Lababex (2015), Trofimovich, et al. (2016)
Turkish	Munro & Derwing (2001)
Taiwanese	Burda (2000)
Ukrainian	Munro & Derwing (2001), Isaacs & Thomson (2013)
Vietnamese	Derwing, et al. (2014), Munro & Derwing (2001)
Japanese	Derwing & Munro (1997), Kang (2010), Kim (2008), Munro & Derwing (2001), Munro, et al. (2006), Saito, Trofimovich, et al. (2016)
Mongolian	Kim (2008),
Farsi	Crowther, Trofimovich, Saito, et al. (2015), Crowther, Trofimovich, Isaacs, et al. (2015), Trofimovich, et al. (2016)
English* (with speakers of L2 German)	O'Brien (2014)
Cantonese	Derwing & Munro (1997), Munro & Derwing (2001), Munro, Derwing, & Morton (2006)
Polish	Derwing & Munro (1997), Munro & Derwing (2001), Isaacs & Thomson (2013)
Nepali	Kang (2010), Kennedy, et al. (2015)
Khmer	Derwing, et al. (2014)

Table 1. L1 of Speaker analysed of Non-Native Speech for Accent and Comprehensibility

As can be seen, one of the most widely used L2 speaking groups were Mandarin speakers with some 13 studies relying either fully or partly on Mandarin accented English. Romance languages (mostly French and Spanish) were analysed just as often, most likely for reasons of proximity as most of these studies focused on Canadian or American L2 speakers of English. It should also be noted that not all researchers specified their speakers' L1 which was the case for Chuang (2010) who only stated the use L2 English International Teaching Assistants. Nearly all studies also used L1 English speaker control groups, mostly from university settings from various departments (e.g.: Anderson-Hsieh & Koehler, 1988; Munro & Derwing, 1995a; Munro & Derwing, 1995b; Burda, 2000, Kennedy & Trofimovich, 2008; Kang, 2010). In total, some 20 different L1 speakers were rated in these studies, with perhaps the greatest variety in Munro and Derwing (2001) in which 12 L1 groups were recorded (Arabic, Cantonese, Japanese, Mandarin, Persian, Polish, Russian, Serbo-Croatian, Spanish, Turkish, Ukrainian and Vietnamese).

Speakers were most often students, as was mentioned, however, in some cases ITA (International Teaching Assistants) were also evaluated for L2 speech (Chuang, 2010; Hsieh, 2011; Kang, 2010). Speakers' gender ratios also seemed even throughout studies and if not, researchers either used only one gender as was the case with Kang (2010) and Kang et al. (2010) who used only male speakers, or gender disparities were made apparent to readers as was done with Trofimovich, Isaacs, Kennedy, & Saito (2016) when the number of female Hindu/Urdu speakers outnumbered male speakers despite the fact that the number of males to females were roughly equal for their other speaker groups (Chinese and Farsi).

Depending on the intent of the study, age varied little from one study to another, however, most did stay in the adult range for participants and most remained in the university years with few extending to senior years. The youngest participants were, for the majority of

these studies, 19 years of age (Trofimovich & Baker, 2006; Derwing & Munro, 1997; Kang, 2010; Munro & Derwing, 2001; Saito, Trofimovich & Isaacs, 2016; Trofimovich & Isaacs, 2012), but even these corresponded to years that individuals would normally start to attend or would already be in university. Nearly all (with the exception again of O'Brien, 2014) were university students either currently in, or previously in, ESL programs. An exception is the 2015 del Puerto, Lacumberri, and Lababex study on high-school level bilingual L1 Spanish/Basque speakers of L2 English, who were all between 14 and 16 years of age. However, their study appears to be one of the only research initiatives specifically linked to accentedness and comprehensibility that used secondary school-aged students and did not employ participants from North America or any other Predominantly English-speaking environment. In short, most speakers were between 20 and 40 years of age at the onset of the studies. Other components of interest to researchers with regards to age have been AOA (age of arrival) (Trofimovich & Baker, 2006; Flege, Munro, MacKay, 1995; Kennedy & Trofimovich, 2013) and LOR (length of residence) (Kennedy & Trofimovich, 2013). This would seem appropriate due to many other studies concerned with "earlier the better" arguments – which, as we will see in a later section of this paper, may offer some evidence to its favor – that seem to give adults little hope for native or nativelike proficiency in their L2.

The number of speakers collected also varied from as few as 3 (Anderson-Hsieh & Koehler, 1988) to over 100 speakers (Saito, Trofimovich & Isaacs, 2016; Trofimovich et al., 2016). Depending on the purpose of the study, speakers either varied in their speech production capacities between beginner, intermediate and advanced (Trofimovich & Baker, 2006; Kennedy & Trofimovich, 2013; Saito, Trofimovich, & Isaacs, 2016; Saito, Webb, Trofimovich & Isaacs, 2016a; Saito, Webb, Trofimovich & Isaacs, 2016b), or represented a more homogenous group of participants.

Listeners

Early on in accentedness and comprehensibility research, listeners were asked to merely rate for accent, comprehensibility, and often added another factor, such as intelligibility, mostly because of Munro and Derwing's (1995a) definition of *intelligibility* (what listeners actually understood, most often evaluated through listener transcriptions of speakers' speech) versus the definition of *comprehensibility* (a rated measure of how well listeners *believed* they understood speakers' speech). Earlier still however, Anderson-Hsieh and Koehler (1988) used listener ratings to determine which speaker rates allowed for most comprehensible speech, in which speech was altered but the accents were mostly homogenous (three L1 Chinese speakers were evaluated). However, from the 1995 Munro and Derwing articles on ward, listeners were mostly used to rate for particular measures of speech (for example segmentals, pitch, optimal speaking speed, lexical, grammatical and prosodic features) to see which of these measures correspond best with either accentedness, comprehensibility, or both. This was because, while accent is often found to be the most salient aspect of foreign speech (Derwing & Munro, 1997; Derwing &

Munro, 2009b), and may contribute to prejudices against L2 speakers (Nguyen, 1993), accent was rarely the root cause of communication breakdown between interlocutors.

In general, it was observed that studies employed roughly equal number of female and male listeners, as was seen with speakers and, just as the number of speaker participants varied, so too did listeners, depending on the goal of the study. For the most part, listener number varied from 18 to over 200, as seen in Anderson-Hsieh and Koehler (1988). Depending on the study, multiple groups with even or uneven numbers of listeners were used. For example, Trofimovich et al. (2012) split their listeners between the aforementioned expert/non-expert groups and used 10 expert listeners and 60 novice listeners, while Munro et al. (2006) used even number of native and non-native listeners. This last study demonstrates that not all listeners used were necessarily native English speakers, as the goal of Munro et al. (2006) was to determine if non-native speakers found other non-native speakers sharing the same L1 more comprehensible than did native speakers of English (according to which the results ascertained negligible significant differences between accentedness and comprehensibility scores between native and non-native listeners). Most studies had between 20 and 40 listeners for rating tasks and these numbers were often related to the number of speakers: often, if there were more speakers, there would be less listeners and vice versa. In cases were the stimuli were very numerous, listeners would be called back for multiple sessions to avoid rater fatigue as was in Munro and Derwing (2001) with features 200 items to be rated, and in Saito, Trofimovich, and Isaacs (2016) which featured 40 files and 11 ratable variables, along with the standard comprehensibility and accentedness rating questions common to nearly all other studies observed here.

In terms of training, listeners did at times receive practice rating exercises in order to habituate themselves to the provided rating scales (see, for example, Munro & Derwing, 2001). However, not all listeners were trained in their tasks. For example, of the studies observed here, Anderson-Hsieh and Koehler (1988), Derwing et al. (1997), Kang (2010), Kennedy et al. (2015) for example did not provide any training to listeners. However, Trofimovich et al. (2012) also trained listeners on a separate day with regards to rating scale use.

Furthermore, because of the general goal of speech research on comprehensibility in particular, all studies focused, in some way, on the communicative capacities of speakers by ways of listener evaluations of speakers' L2 speech (see Table 2).

Raters	**Research Authors**
Native expert	Crowther, Trofimovich, Saito, et al. (2015), Crowther, Trofimovich, Isaacs, et al. (2015), Isaacs & Thomson (2013), Kennedy & Trofimovich (2008), Saito, Trofimovich, et al. (2016), Trofimovich & Isaacs (2012)
Non-native expert	Del Puerto et al. (2015)
Native Novice	Chuang (2010), Crowther, Trofimovich, Isaacs, et al. (2015), Isaacs & Thomson (2013), Kennedy & Trofimovich (2008), Del Puerto, et al. (2015), Saito, Trofimovich, et al. (2016), Trofimovich & Isaacs (2012),
Non-Native Novice	Chuang (2010), Kennedy, et al. (2015), Munro, et al. (2006)

Table 2. Rater Types for L2 English Speech Production Tasks

These listeners, as seen in Table 2, vary from being either L1 English speakers, L2 English speakers, novice – meaning no training in L2 English speech as defined by Isaacs & Trofimovich (2012) – or expert – defined as have training or a number of years teaching L2 English (Kennedy & Trofimovich, 2008, Trofimovich & Isaacs, 2012). It should be noted that while all studies used raters, only those seen below focused on, or mentioned using expert vs. non-expert ratings of L2 English speech. In general, many studies found the rating results between expert and non-expert listeners were similar, though Isaacs & Thomson (2013) found greater intergroup reliability among expert raters than among novice raters. The expert/non-expert listener dichotomy is also a reflection of the fact that, for all studies in Table 2, researchers focused on teaching techniques to further advance L2 speech, and in all expert cases, listeners were either experienced teachers or learning to become ESL teachers in the future.

The results obtained from teacher-raters was often intended to lead to better classroom-based innovations in L2 speech. For instance, in their recent study that looked that non-expert vs. expert ratings of L2 speakers of English on comprehensibility, Saito, Webb, Trofimovich, and Isaacs (2016a) demonstrated that teachers would often show better understanding of such speech measures as grammatical complexity, and were also more likely to use a greater variety of speech measures when it came to rating L2 speech than did non-expert raters. The study demonstrated, not only what kinds of elements teachers needed to focus on to improve speaker comprehensibility, but also on more integrative approaches to teaching oral proficiency, through focus on fluency, lexical and grammatical features, rather than just pronunciation, which is what contemporary L2 teaching textbooks appear to focus on (Saito, Trofimovich & Isaacs, 2016).

Stimuli Types

Rated stimuli types ranged greatly between groups, but in general could be narrowed down to broader categories: read stimuli, or extemporaneous narratives, with a full list in Table 3. Interestingly, the great majority of researchers in North America that used extemporaneous elicitations used the 2008 Munro et al. suitcase story, about 2 individuals carrying identical suit-cases that then bump into each other and accidentally switch lugged, which may have been more appropriate for adult speakers. This could explain why the adolescent speakers from Spain (del Puerto et al., 2015) used a Mayer (1969) picture story called *Frog, where are you?* This story could have been better suited for younger speakers.

Stimuli types also depended upon the goal of the rating tasks, as demonstrated in Table 3.

Stimuli	**Studies**
Read: Ex: Paragraph or True/False statements	Anderson-Hsieh & Koehler (1988), Burda (2000), Derwing, et al. (2014), Kennedy & Trofimovich (2008), Kraut & Wulff (2013), Munro & Derwing (1995b), Munro & Derwng (2001), Munro, et al. (2006), Trofimovich & Isaacs (2012)
Read: Ex: Meaningful with no context	Kennedy & Trofimovich (2008)
Read: Ex: Not meaningful	Kennedy & Trofimovich (2008),
Extemporaneous: solitary (ex: picture stories)	Crowther, Trofimovich, Saito, et al. (2015), Derwing & Munro (1997), Isaacs & Trofimovich (2012), Isaacs, &

	Thomson (2013), Kim (2008), Munro & Derwing (1995a), O'Brien (2014), Saito, Trofimovich, et al. (2016), Trofimovich & Isaacs (2012), Trofimovich et al. (2016)
Extemporaneous: partner or group (ex: language exam, interview)	Hsieh (2011), Kang (2010), Kang, et al. (2010), Kennedy & Trofimovich (2013), Kennedy, et al. (2015)
Extemporaneous: delayed repetition	Trofimovich & Baker (2006)

Table 3. Recorded Stimuli for Listener Rating Exercises

Although today, results from accentedness and comprehensibility rely on numerous factors to determine the similarities and differences between those two groups, in earlier studies, in Table 3, researchers chose reading tasks for speakers that would eliminate most grammatical or lexical variances so that listeners focused mostly on phonological aspects of speech. Munro and Derwing's 1995 study was in response to the idea that improving pronunciation is directly correlated with improved comprehensibility. However, their results, and the results of a subsequent 1995 study indicated that listener scores for accentedness were not adequate indicators for comprehensibility assessment, and that listeners pay particular attention to traits such as speech speed when evaluating comprehensibility. As more researchers began to have various dimensions of speech rated in terms of accent and comprehensibility, studies leaned towards the use of extemporaneous narratives in which speech can differ from one level of oral proficiency, and indeed, from one accent group to another.

Stimuli length also varied between studies. In most cases, stimuli were shorter, between 20 to 60 seconds in length such as Trofimovich et al. (2016) who used 30 second sound bites and Kang et al. (2010) who used minute long clips however, some studies such as Saito, Trofimovich, et al. (2016) used full length recordings for their rating tasks. Depending on the demands of the study, certain recordings were "cleaned up", that is to say, removed of false starts and hesitation markers (for example: Trofimovich & Isaacs, 2012; Isaacs & Thomson, 2013; Crowther, Trofimovich, Saito, & Isaacs, 2015; O'Brien, 2014). Some studies went further still. In Baker and Trofimovich (2006) the stimuli (240 sentences) were further treated to muffle content but preserve suprasegmental features that listeners then used to rate accentedness. This was so because accentedness was the key feature, rather than the combination of accent and comprehensibility. Some other studies such as Crowther et al. (2015), Trofimovich et al. (2016), and Saito, Trofimovich, et al. (2016) used transcriptions of audio files as well as voice recordings to rate particular measures of speech such as lexical and grammatical features.

Rating Scale Types

Ratings scales were common tools among nearly all observed studies in this report (Kim (2008) employed a questionnaire to rate speakers). It should be mentioned as well that not all rating scales were used exclusively to measure accentedness, comprehensibility and their features. Some studies like O'Brien (2014) employed a 4-point scale first for self-assessment ratings, while Crowther et al. (2015), and Trofimovich et al. (2016) used a 9-point scale to assess

how easy or difficult listeners found their rating tasks to be, while the actual measurements of accent and comprehensibility used 9-point and continuous sliding scales respectively.

As can be seen in Table 4, the most common rating scale types were 9-point scales.

Rating Scale	**Studies**
4-point scale	O'Brien (2014)
5-point scale	Anderson-Hsieh & Koehler (1988) Isaacs & Thomson (2013), Sereno, Lammars & Jongman (2016)
6-point scale	Saito, Trofimovich & Isaacs (2015), Isaacs & Trofimovich (6-point)
7-point scale	Burda (2000), Kang (2010), Kang et al. (2010), Kraut & Wulff (2013)
9-point scale	Munro & Derwing (1995a), Munro & Derwing (2001),Trofimovich & Baker (2006), Kennedy & Trofimovich (2008), Munro et al. (2006), Isaacs & Trofimovich (2012), Trofimovich & Isaacs (2012), Isaacs & Thomson (2013) , Crowther, Trofimovich, Saito et al. (2015), Isaacs et al. (2014), O'Brien (2014), Kennedy, Foote, Kurtz, & dos Santos Buss (2015), del Puerto et al. (2015), Isaacs & Trofimovich (2016),
Continuous sliding scale	Crowther, Trofimovich, Saito, et al. (2015), Trofimovich et al. (2016), Saito, Trofimovich, et al. (2016), Saito, Trofimovich, & Isaacs (2017), Saito, Webb, et al. (2016a), Saito, Webb, et al. (2016b)

Table 4. Rating Scale Types for Accentedness and Comprehensibility Studies

This is unsurprising since an often cited 1999 study by Southwood and Flege observed that 7-point scales were more prone to ceiling effects among listeners who were rating longer elicitations, and suggested that, to curtail these effects, 9-point or even 11-point scales would be better suited for rating tasks. As most studies observed here used longer and often extemporaneous recording excerpts, it appears to make sense that they would employ 9-point scales for their rating tasks.

However, in a later study on rating scales on L2 pronunciation, Isaacs and Thomson (2013) noted in their findings that, depending on the exercise type, not only did lower point scales (such as 5-point scales) not produce a ceiling effect, results showed no significant difference on any of the analysed dependent variable measures and rating scale lengths. Listeners also suggested that the 9-point scale might have been too long for the exercise and that the 5-point scale was better suited for the exercise (Isaacs & Thomas, 2013). This led to them concluding that 9 and 11-point scales were not always optimal rating tools for all circumstances. Isaacs, Trofimovich, et al. (2016) also looked at comprehensibility rating scales to be used in L2 English university classrooms to evaluate L2 speech, and finally concluded upon a 6-point scale for raters, and many studies today (such as: Crowther, Trofimovich, Saito et al., 2015; Trofimovich et al., 2016; Saito, Trofimovich, & Isaacs, 2015; Saito, Webb, Trofimovich, & Isaacs, 2015) use continuous or free-moving siding scales that measure from 0 to 1000, but do not show any numeric values to listeners.

Results Found in Recent Research on Accentedness and Comprehensibility

The results of these studies can be grouped into approximately four categories depending on the purpose of the studies. Studies looked at age related issues surrounding accent and

comprehensibility, whether or not experience with L2 speaking influences ratings, and L1 listener attitudes towards L2 speakers. The majority of studies observed here were on identifying features of accent and comprehensibility in order to find where the two differ. This first part looks at any studies regarding age, followed by listener attitudes and finally at the identified features of accent and comprehensibility.

Speaker Age Effects

Flege et al. (1995) concluded that, among other features such as gender and relative use of L2, both age of arrival (AOA) and length of residence (LOR) affected speakers' perceived accents. This prompted further studies which did not always yield the same results. Trofimovich and Baker (2006) looked at AOA versus LOR regarding listener ratings for accentedness. Speakers were split into three groups (beginner, intermediate and advanced speakers of English), though all speakers were at the time of the study – and had been upon arrival to the US – 18 or over the age of 18. It was found that LOR had no significant effect on suprasegmental proficiency, but AOA did. Saito, Trofimovich, and Isaacs (2017) chose speakers with varying AOA because of the belief that L2 speakers who arrived in their L2 environment at an earlier age had better segmental and suprasegmental proficiency. Kennedy and Trofimovich (2013) demonstrated that there was no net difference in ratings of comprehensibility and accentedness scores for first and final semester non-native university students, showing once again that LOR had no impact on ratings while AOA might.

Another interpretation of age-related issues was done by Burda (2000) who looked at listener ages of 72 L1 English speakers. The study assumed that it is possible that older listeners were more tolerant of L2 speech. Of the three age groups (20-39, 40-59 and 60 and older) older listeners were in fact found to have more difficulty understanding L2 speech, though there appeared to be no net significance between listener age groups and measures of accentedness, comprehensibility and intelligibility.

L1 Listener Attitudes Towards L2 Speakers

Anderson-Hsieh and Koehler's (1988) study on non-native accent effects on listener ratings of L2 speech suggested that raters with more positive attitudes towards foreigners tended to rate accentedness less harshly even when speech was faster (which otherwise often lead to harsher ratings for both accent and comprehensibility). Nguyen (1993) suggested that while some researchers have tried to focus more on comprehensibility rather than accent eradication, accented speech comes with many stigmas in society, which would reiterate the previous studies results, and could explain why some studies like Munro and Derwing (1995a) found that while listeners were very accurate in their transcriptions of L2 speakers (indicating high intelligibility), listeners still rated heavily accented speakers more harshly for comprehensibility as well. Using a foreign accent questionnaire, Kim (2008) demonstrated a bias even from L2 speakers towards L2 international Teaching Assistants (TA). The ESL speakers claimed to fear picking up their L2 English teacher's foreign pronunciation.

In a general study on accent, Derwing and Munro (2009a) noted that while there appears to be no correlation between comprehensibility and degree of accentedness, an accent could lead to loss of intelligibility as well as discrimination towards to individual. Chuang's (2010) study on attitudes towards L2 speaking international TAs, demonstrated that for the most part, ITAs were not negatively perceived by their students, but that harsher ratings of accentedness did come from students with negative attitudes towards their foreign accented TAs. Ballard (2013) also found that accent correlated both with comprehensibility and teacher acceptability, and concluded that students should get greater exposure to various accents.

Uses of Expert and Novice Raters

Studies looking at whether "expert" or "novice" listeners make for more accurate ratings have come to various conclusions about whether there are any differences between groups. While some studies observed here, concluded that experienced judges were significantly more accurate in their ratings of L2 speech (Kennedy & Trofimovich, 2008), many generally concluded that there was no net difference between ratings by expert, often described as individuals with English L2 teaching experience (Kennedy & Trofimovich, 2008; Kim, 2008; Isaacs & Trofimovich, 2012), and non-expert rater, and even the 2008 Kennedy et al. study concluded no net difference for comprehensibility ratings between expert and non-expert raters. This was also found to be true, for example, in Isaacs and Thomson (2013) where there appeared to be a net difference in response time between expert and non-expert listeners (experienced raters took longer to finish the tasks), but there remained no significant difference between rating results.

While in general, studies demonstrated no difference in rating results as seen above, at times, experts could have an easier time understanding L2 speakers, even if they themselves were L2 speakers as was the case with ratings of Spanish/Catalan bilingual speaker in Spain (del Puerto et al., 2015). The point that expert raters either rated differently or in a more detailed manner than novice raters was suggested in the finds of many studies (Kennedy & Trofimovich, 2008; Isaacs & Trofimovich, 2012; Crowther, Trofimovich, Saito, et al., 2015). Isaacs et al. (2012), for example selected both expert and novice listeners, so that the expert raters could further elaborate on grammatical structures in their ratings.

Features of Accentedness and Comprehensibility

As was mentioned before, the bulk of the studies observed for this report looked at how and to what extent comprehensibility and accentedness were related, as well as what features of each separated to two. An earlier study of accentedness and comprehensibility by Anderson-Hsieh and Koehler (1988) on L2 speaking rates and comprehension found that heavily accented speech specifically with heavy segmental deviances affected comprehensibility as did faster speech of L2 speakers. As was mentioned, their results were somewhat influenced by rater attitudes towards L2 accented speech, for which subsequent studies noted that comprehensibility itself should be a measure of how much listeners thought they understood the L2 speech (called

"perceived comprehensibility" by Munro and Derwing,1995a) and not a measure of what they actually understood.

The 1995a Munro and Derwing article looked at differences between accent and comprehensibility, as often the two were seen to be completely related. Their study, and their subsequent 1995b study suggested that while the two did correlate with one another (as was again found by Burda, 2000), a strong foreign accent did not impede comprehensibility, but even though listeners might take slightly longer to rate a speaker, this may not be linked to speaker`s accentedness. A 1997 extension of the Munro and Derwing studies looked at grammatical features and phonemic features with regards to accentedness and comprehensibility (again with L1 English raters), and concluded that in order to aid L2 speaking in improving their oral proficiency, teachers should also focus on teaching grammatical and prosodic features as opposed to only phonemic features. These studies would lead to the idea that while ratings for both accent and comprehensibility did correlate, they were independent features, which would be reiterated in a Derwing and Munro (2009a) study on L2 speech in the work place, where comprehensibility was an important factor in listener preferences in L2 interlocutors, but accentedness was less important.

As was mentioned grammatical and prosodic features did correlate with comprehensibility specifically. Other features that have been found to have an effect on both accentedness and comprehensibility were speaker rates (Munro & Derwing, 2001) and semantic intelligibility as seen in Kennedy and Trofimovich (2008) where statements that were intentionally less intelligible were perceived as less comprehensible and more accented than intentionally more intelligible statements (ex: recordings of true/false statements). The Kennedy and Trofimovich (2008) as well as a Munro (2006) study also shifted attention away from comprehensibility as a measure of 'perception of meaning', to a measure of 'ease of understanding'. A further study by Kang (2010) on international Teaching Assistants (ITA) demonstrated that various types of suprasegmentals effect comprehensibility and accentedness independently: accent pitch and word stress were associated with accentedness and speaking rates were associated with comprehensibility.

Further linguistic measures were looked at to see which correspond to accentedness and comprehensibility in Isaacs and Trofimovich (2012). They selected 19 different speech measures from 4 categories – phonology (6 features: segmental error ratio, syllable structure error ratio, word stress error ration, vowel reduction ratio, pitch contour and pitch range), fluency (6 features: total number of pauses, pause error ratio or inappropriate pauses, total number of unfilled pauses, repetition and self-correction, pruned syllables per second, and mean length of run), linguistic resources (4 features: grammatical accuracy, lexical error ratio, token frequency and type frequency) and discourse features (3 features: story breadth, story cohesion and story depth) – of which 18 correlated with comprehensibility ratings, showing that listeners, especially expert listeners, used many linguistic measures such as grammar, vocabulary, fluency in L2, word stress discourse structure, context and familiarity with the speaker's L1, which judging comprehensibility.

A subsequent study on the same 19 measures found that 8 measures significantly correlated with both accentedness and comprehensibility and that raters tended to focus on segmental accuracy such as syllable errors with regards to accentedness, but that these features were less important for comprehensibility. Grammatical and certain lexical errors were linked to comprehensibility ratings. Most studies seem to demonstrate that accentedness ratings are linked to pronunciation, while lexicogrammar as well as phonological features are linked to comprehensibility (Crowther, Trofimovich, Saito, et al., 2015; O'Brien, 2014; Trofimvich & Isaacs, 2012; Saito, Trofimovich & Isaacs, 2015; Crowther, Trofimovich, Isaacs, et al., 2015; Saito, Trofimovich & Isaacs, 2015). Further studies focused specifically on one or more of these measures as was the case with Saito et al. (2016a) and Saito et al. (2016b) that looked at lexical features that were most associated with comprehensibility depending on the level of proficiency of each speaker. Comprehensibility for beginner and intermediate speakers was found to be associated with the fluency and accurate use of vocabulary, and for intermediate and advanced L2 English speakers, morphological accuracy, and lexica complexity were associated with comprehensibility scores.

Languages themselves were also linked to comprehensibility in some studies. Crowther et al. (2015) found that certain languages such as Farsi and Hindu/Urdu were rated as more comprehensible and accented to L1 listeners than Chinese groups, with Hindu/Urdu being rated as more comprehensible than Farsi, and Chinese being rated as more accented than Farsi for example. Furthermore, while being familiar with various L1s did (but not always, see Isaacs & Thomson, 2013) demonstrate more consistent, detailed or accurate ratings from listeners (Kennedy & Trofimovich, 2008; Isaacs & Trofimovich, 2012; Crowther, Trofimovich, Saito, et al., 2015), sharing an L1 with the speaker did not (Burda, 2000; Munro, Derwing, & Morton, 2006; Kim, 2008).

Summary and Concluding Thoughts

This report has looked at the differences and similarities in accentedness and comprehensibility by ways of rating L2 speech. The purpose of most of these studies was to observe if the two terms were independent of each other. Speakers for these studies were taken from a wide variety of L1 groups, though the most numerous were found to be Mandarin and Romance language speakers. Listeners, or raters, were most often L1 speakers of the target language (almost exclusively English), and could be either expert (individuals with L2 English teaching experience) or novice, with no teaching experience. While the two were often found to be correlated to a certain extent, heavily accented speech was not found to be necessarily incomprehensible. Phonemic features were often found to be related to accentedness ratings. Comprehensibility was most often associated with both pronunciation, such as prosodic features and speaking rate, and so-called linguistic resources (Isaacs & Trofimovich, 2012), such as grammatical complexity and vocabulary.

It was interesting to note that while there are many languages being taught in schools and through private programs, out of the 38 studies conducted with regards to accentedness and

comprehensibility, all except 1, have focused on English as the target L2. Globally, these studies actively advocate taking a step from monolingual approaches to accentedness and comprehensibility. However, by focusing almost exclusively on English, and by using it as a de facto generalization tool for studies on accentedness and comprehensibility, many of these researchers have, perhaps unintentionally, affirmed the dominance of English. Indeed, the fact that almost none of the papers had explicitly stated that they were using *English* acc/comp research (as opposed to general acc/comp research), seems to indicate that English, in this field, is treated as a lingua franca that can be used as a template for all languages, or, on a more sinister level, that no other language is relevant or necessary enough to research. It also excludes so-called 'native' English speakers from testing, as the focus is predominately on the accentedness and comprehensibility of English language learners, rather than English speakers learning other languages. By extension, we, as language education researchers in the SLA field are, without even noticing, perpetuation the myth of English as a global language which is simply, in Kubota's (2016) words, "reinforcing the hegemony of English monolingualism", both in academia, and the language classroom. If we are to further expand on acc/comp research in the future, it would be more prudent to further investigate other L2 languages and compare these results with those found by English accentedness and comprehensibility researchers. Only then could we begin generalizing our results on L2 research, while still maintaining that mantra that all languages be equally important, and all L2 learners be equally valued.

In the following appendix, a number of studies of accent and comprehensibility have been summarized. Specifically, 26 have been selected because they pertain to the results seen above and are found to be the most relevant studies on accentedness and comprehensibility for the sake of this report.

Appendix 1

1) *1988 Janet Anderson-Hsieh, Kenneth Koehler: The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension*

In this study, three L1 Chinese and 1 L1 English control recorded read passaged in various speeds which speakers modified themselves because speech synthesizers were not available to the researchers. L1 English listeners were given 6 multiple choice questions to test comprehensibility as well as a 5-point scale to rate accentedness and comprehensibility, though they were not informed that some of the speakers would be non-native, nor did they receive any particular training for the rating tasks. After listeners rated speakers' speech for both accent and comprehensibility, it was found that comprehension scores were significantly higher for the native passages than for the non-native passages and in particular, the scores were significantly higher at the regular rate than at the fastest rate for all speakers. It was also found that the increase in speaking rate from the regular to the fast rate resulted in a greater decrease in comprehension more so for the most heavily accented speaker than for the other speakers, suggesting that speaking rate is more critical for the comprehension of heavily accented speech, and that prosodic errors affected comprehension more than did segmental deviance.

2) *1995 Murray Munro, Tracy Derwing: Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners*

In this study on the relationship between accentedness, perceived comprehensibility and intelligibility in the speech of L2 leaners, listeners were given recordings of L2 university students as well from an English L1 control speaker. All speakers were recorded in a sound-treated room. Listeners, who were L1 English speakers, had taken some linguistics or teaching methodology courses, and had a basic knowledge of phonetics were asked to first rate speakers for accentedness and perceived comprehensibility, and then to transcribe as accurately as possible what they had heard. Results indicated that the strength of the accent was found to correlate with comprehensibility ratings, but having a strong L2 accent did not impede comprehension according to the ratings. Comprehensibility score were also less harsh than accentedness scores. This seemed to suggest that accentedness was a poorer measure of comprehensibility indicating that the two might be independent of each other, which was in opposition with the contemporary idea that improving accentedness directly lead to grater comprehensibility. The study also suggested that further research should focus on grammatical and lexical features of speech and how they correlate with accentedness and comprehensibility.

3) *1995 Processing Time, Accent and Comprehensibility in the Perception of Native and Foreign-Accented Speech*

This was the second Munro and Derwing study on accentedness and comprehensibility that continued from the previous study. L2 speakers recorded read true/false statements with a mean length of 5.9 words per person. This article also looked at processing time as well as ratings for accent and comprehensibility and speech transcriptions. The results indicated that listeners made

more errors while transcribing sentences produced by non-native speakers than by native speakers, and sometimes due to accent, comprehension was fully blocked. The Mandarin speakers' utterances took significantly longer to verify than did native-speakers, and while accentedness and comprehensibility ratings were linked, there were cases were listeners rated accent much more harshly to completely comprehensible utterances. This suggested that even shorter utterances could be rated as comprehensible even when speakers are heavily accented. There was also no relationship between response time and accentedness, though there was for comprehensibility and response time.

4) *1997 Tracy Derwing, Murray Munro, Accent Intelligibility and comprehensibility: Evidence from Four L1s,*

Also an extension of previous study on comprehensibility and accentedness and intelligibility now with varying speaker L1 types. High proficiency L2 speakers from 4 different L1 backgrounds were recorded for accentedness and comprehensibility ratings. Again, transcriptions were made of each recording by listeners, and this time, listeners also rated for grammatical errors, phonemic errors and speaker prosody. Results indicated that being a high proficiency learner did not affect the relationships between intelligibility comprehensibility and accentedness, however individual features of each group, such as grammatical and phonemic errors and prosody still differed. Accentedness was still rated more harshly than was comprehensibility, but accent and comprehensibility scores were related but not dependent of each other. It was also found that even though some aspects of accent were more salient, they did not necessarily interfere with intelligibility. This suggested that improving other aspects of speech such as grammar and prosody could lead to improved comprehensibility more so than improved phonemic features alone.

5) *2000 Angela Burda Language and Age Variables Affecting Measures of Intelligibility, Comprehensibility and Accentedness*

The purpose of the study was to look at if age or native language of the speaker effect listeners' measures of intelligibility, comprehensibility and accentedness. Listeners of various age groups Listeners were in various age groups rated three speakers. Results suggested that accentedness and comprehensibility correlated (as ratings for comprehensibility increased ratings for accentedness decreased), and no age effects existed for either comprehensibility scores or accentedness scores, however, older listeners did have greater difficulty in understanding accented speech, especially at the word and sentence levels, specifically in this study older listeners found Spanish speakers the most difficult to understand

6) *2001 Murray Munro, Tracy Derwing, Modeling perceptions of the accentedness and comprehensibility of L2 Speech: the Role of Speaking Rate*

This study also looks at speaking rate as did the Anderson-Hsieh and Koehler (1988) study due to the fact that the researchers wanted to add an element outside of segmental and prosodic

features, and because previous research has also looked at speaking rates. The study was split into two parts. For the first part, L2 speakers recorded sentences that were rated by L1 speakers of English (who had no training in the rating procedure). Both accentedness and comprehensibility rating results correlated with speaking rates (speaker speech was not treated in this part of the study). The second part of the study took the same sentences produced by the same L2 speakers and sped half of them up by 10% and slowed down the other half by 10%. This time the rating results from listeners indicated that listeners preferred the faster speech rates which may be due to the fact that speaker`s original speech rates were slower than L1 speech to begin with. Overall it was suggested that speaker rate contributes to but accentedness and comprehensibility, and that it does so independently of phonological errors.

7) *2006 The Mutual intelligibility of L2 speech: Murray Munro, Tracy Derwing, Susan Morton*
The general premise of the research looked at if non-native speakers find non-native English speech to be less accent, and more comprehensible than Native-speakers of English do. The speakers and the stimuli were the same was for Munro and Derwing (2001). Listeners were, as suggested, but L1 and L2 speakers of English who performed evaluations on accentedness and comprehensibility and found that having a similar L1 did not necessarily aid in comprehension, and while it did appear that L1 Cantonese listeners did rate L1 Cantonese speakers are more comprehensibile, intelligibility results indicated that they were no more comprehensible to the Cantonese listeners than were any other language group. Overall, it was found that native and non-native listeners rated non-native listeners similarly, suggesting that similar rating processes are used among both groups.

8) *2008 Intelligibility, Comprehensibility and Accentedness of L2 Speech: The Role of Listener Experience and Semantic Context Sara Kennedy, Pavel Trofimovich*
Study wanted to address two questions: first it wanted to see if ratings for accentedness and comprehensibility vary according to the degree of meaning, and second, do more experienced raters rate different from novice raters with respect to degrees of semantic context. L2 speakers were recorded reading sentences that varied in how meaningful they were. First they recorded true/false statements that were the most meaningful, then they recorded a semantically meaning sentence with no context provided, and finally a sentence that was grammatically correct, but had no meaning or context. Listeners then performed two rating tasks and a vocabulary knowledge test. The results indicated that semantic context had an effect: if something was more difficult to understand it was also rated as more difficult, but also as more accented however, both experienced and inexperienced listeners understood L1 English speech that was used as a control. However, experienced listeners were significantly more accurate than inexperienced listeners in ratings and vocabulary test result. The researchers concluded that this may be due to experienced listeners having more knowledge of how L2 speakers pronunciation differs from L1 English. However, experienced listeners did not rate L2 speakers as more comprehensible than inexperienced speakers, and accentedness was rated the same way by experienced and

inexperienced listeners, suggesting that both listener groups used similar criteria to evaluate accent.

9) *2006 Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech, Pavel Trofimovich, Wendy Baker*

This study looked at 5 suprasegmentals from 30 Korean learners of English to see how they contribute to accent, and in particular, if length of residence affects mastery of suprasegmentals. The speakers were grouped into three categories of language proficiency: beginner, intermediate and advanced. Each language group often corresponded to the number of years they lived in their L2 language environment. According to the results of the L1 speakers of English who rated various factors of foreign accent (stress time, peak alignment, speech rate and pause frequency), only stress-time was related to length of residence, and speech rate, pause frequency and pause duration were linked to age of arrival, indicating that more suprasegmental features were linked to age of arrival than the duration of their stay in their L2 language environment. Peak alignment was not related to either age of arrival or length of residence.

10) *2008 Accentedness, comprehensibility, Intelligibility and Interpretability of NNESTs Taesung Kim*

This study looked at the ratings for accentedness and comprehensibility (which was defined here as listener's perceived level of difficulty in understanding L2 speech) of non-native English speaking teachers by other ESL students. First, L2 listeners were asked to fill out a questionnaire on their attitudes towards foreign TAs. Then, L2 listener rated for accent and comprehensibility, and results appear to indicate that while no stimuli were rated as more or less comprehensible, attitudes towards foreign TAs were such that L2 listeners believed that they should be taught by native speakers of English. Accent also appeared to having an effect on perceived comprehensibility because foreign-accented speech was believed to be difficult to understand, even though results suggested that this was not the case.

11) *2010 Relative salience of suprasegmental features on judgement of L2 comprehensibility and accentedness, Okim Kang*

This study focused on how suprasegmentals affect listeners judgements of L2 speakers accented speech. Specifically, the study focused on the speech of L2 accented foreign teaching assistants in the United States in an attempt to isolate the most salient suprasegmentals that affect comprehensibility ratings. It was found that they do in fact contribute independently to listeners' ratings: accent was most often associated with pitch range and word stress and comprehensibility was associated with speaking rates, which demonstrated a linear relationship. TA speech was also often rated as monotonous and flat.

12) *2010 Suprasegmental Measures of Accentedness and Judgements of Language Learner Proficiency in Oral English Okim Kang, Don Rubin, Lucy Pickering*

Using listener ratings as well as measuring instruments to avoid rater bias, this study aimed to identify which features of speech are associated with accentedness and comprehensibility to see which features should be taught in language classrooms to help improve L2 comprehensibility. Speakers were L2 English learner from a variety of language backgrounds were recorded, and files were transcribed as well as converted to .wav formats. Listeners who were all L1 English speakers, rated the L2 speakers were a number of features including pronunciation, grammatical accuracy, vocabulary, speech rate, organization and for the appropriateness of their responses to the task. Results suggested that fluency is an intonational phenomenon, rising tones are associated with comprehensibility and proficiency ratings. The researchers also suggest that in terms of pronunciation instruction, the enhancement of comprehensibility.

13) *2012 Deconstructing Comprehensibility: Identifying the Linguistics Influences on Listeners' L2 Comprehensibility Ratings, Talia Isaacs, Pavel Trofimovich*

The study examined which particular measures of L2 speech contributed to comprehensibility, and was interested in what linguistic measures were most strongly associated with non-expert ratings of comprehensibility and which are associated with expert ratings, and finally, which features distinguish beginner, intermedia and high proficiency L2 levels. All speakers were French L1 speakers which recorded extemporaneous narratives. These were rated for 19 separate linguistics measures (6 features of phonology, 6 features of fluency, 4 features of linguistic resources and 4 features of discourse). Some correlations were found for several measures in each of the conceptual categories of phonology (word stress error ration, vowel reduction ratio) fluency, linguistic resources. Expert raters paid specific attention to grammatical features, and most commented on generic errors, though some pointed to verb errors and sometimes pronoun and preposition errors, and all together, this suggested that experienced listeners rely on many factors when judging L2 comprehensibility such as grammar, vocabulary, fluency in L2 speech. Of the 19 speech measures, 18 significantly correlated with mean L2 comprehensibility ratings, and there appeared to be a relationship between comprehensibility and word stress.

14) *2012 Disentangling accent from comprehensibility, Pavel Trofimovich, Talia Isaacs*

Accent and comprehensibility are portrayed in society as going hand in hand and can create problems for L2 speakers, therefore, this study intends to find which aspects of language belong to accent, and which to comprehensibility. French L1 speakers of English recorded extemporaneous and reading tasks. Each task was normalized and only 23 to 26 seconds were used as stimuli to be rated. The study isolated 19 different language measures for analysis as seen in Isaacs and Trofimovich (2012). L1 English listeners rated for these features and the study concluded that both Accentedness and comprehensibility correlated strongly with 8 of the 19 measures, of those, 6 were strongly associated with both accent and comprehensibility (word stress, rhythm, mean length of run, type frequency, token frequency story breadth), and 2 were unique to accentedness (segmental errors) and comprehensibility (grammatical accurate). Through subsequent regression analysis results, 4 measures were isolated for accentedness (word

stress, rhythm, type frequency, segmental errors) and 4 for comprehensibility (word stress, rhythm, type frequency, grammatical accuracy. Overall, pronunciation features seem to be associated more with Accentedness while grammatical and lexical features were more associated with comprehensibility ratings.

15) *2013 Rater Experience, Rating Scale Length, and Judgements of L2 Pronunciation: Revisiting Research Conventions, Talia Isaacs, Ron, Thomson*

The study examined the effects of ratings scale length and rater experience on listener ratings of L2 speech, with particular interest in 9-point scales, which were thought to be optimal for accentedness and comprehensibility ratings. The speakers were all L2 newcomers to Canada from two different language backgrounds. The tools for the recordings were taken from Munro et al. 2009, and were given to L1 English listeners, half of whom were expert raters and half of whom were novice raters. Listeners were either given a 5-point scale or a 9-point scale, and expert listeners were measured and found to take longer in finishing the exercise than novice raters. There was greater consensus between expert raters, but that overall, there was no net difference between expert and novice rater results. In terms of the scale preferences, some indicated that the 9-point scale was too long for some rating tasks suggesting that 9-point scales were not always optimal for accentedness and comprehensibility tasks. Results seemed to indicate that raters had trouble differentiating between scale steps particularly from the mid range of the scale, which was more apparent with the 9-point scale.

16) *2013 First and final-semester non-native students in an English-medium university: judgments of their speech by university peers, Sara Kennedy, Pavel Trofimovich*

In this study, L2 speakers were rated by two groups of listeners for accentedness, comprehensibility, fluency and communicative effectiveness. In particular, the study looked at whether or not L2 speaking students improve their oral proficiency over time, and if human resource management students, serving as listeners, would rate L2 speakers differently from other rater groups. The results suggested that, for the first question, there was no significant difference between first and final semester L2 speakers. This may suggest that length of residence does not have an effect of speaker proficiency. For the second question, it was found that human resource management students rated accentedness more harshly than did other groups, but the other groups were harsher for comprehensibility than were human resource management listeners. No differences were found between rater groups in their ratings of fluency and communicative effectiveness.

17) *2015 Second Language Comprehensibility Revisited: Investigating the effects of Learner Background Dustin Crowther, Pavel Trofimovich, Kazuya Saito, Talia Isaacs*

This study focuses on L1 effects on listener ratings of comprehensibility, and accentedness in L1 speech. The objectives of the study were to clarify which features of L2 speech contribute to listener perceptions of accentedness and comprehensibility, and whether these features differ as a

function of speakers' L1 backgrounds. Speakers, who were otherwise similar to each other except for having 3 different language backgrounds (Farsi, Hindu/Urdu and Chinese), were selected for analysis by L1 Listeners of English. They recorded extemporaneous speech tasks that listeners then evaluated based on 10 rater categories using a continuous sliding scale. Listeners also rated the exercises themselves so assess to extent to which they understood the categories they rated. Results indicated that for comprehensibility and accentedness. Accentedness was linked exclusively to pronunciation (and its measures), while comprehensibility was linked to lexicogrammar and pronunciation together. Furthermore, comprehensibility was associated with pronunciation for the Chinese group, with lexicogrammar for the Hindu-Urdu groups and with neither factor for the Farsi group. Accentedness was linked to all groups, from segmental issues associated with the Chinese group, segmental issues and intonational and word stress issues associated with the Hindu/Urdu group and segmental and word stress issues for associated with the Farsi group. However, for comprehensibility, only segmental issues were linked specifically to Chinese speakers, whereas lexicogrammar issues were linked to Hindu/Urdu speakers, and nothing was associated specifically with Farsi.

18) *2016 Flawed self-assessment: Investigating self- and other perception of second language speech, Pavel Trofimovich, Talia Isaacs, Sara Kennedy, Kazuya Saito, Dustin Crowther*

This study used accentedness and comprehensibility to observe how speakers' self assessments were compared to L1 English listeners' assessments of L2 speech. In the first half of the study, all speakers self-rated, while in the second part, a portion of the speakers were randomly selected to be evaluated by L1 English speakers on a continuous sliding scale. The results demonstrate that speaker self-ratings related little to their actual performance and that they tended to either over or underestimated their performance, specifically, lower proficiency speakers overestimated themselves, and higher proficiency speakers underestimated themselves. For comprehensibility and accentedness in particular, speakers had discrepancies in judgments compared to L1 English speakers again with people at lower end overestimating themselves and people at upper end underestimating their abilities, however, language backgrounds could not account for these discrepancies. The results of the second part of the study suggest that listeners' judgements were more accurate than speakers' judgements, and for both accent and comprehensibility, discrepancies in self- versus listener assessment were associated with several segmental and suprasegmental features of L2 speech (segmental accuracy, word stress, rhythm, intonation, speech rate) but not with aspects of lexis, grammar, and discourse. In terms of language discrepancies, and the weakest group (Chinese) was more overconfident than Romance and Farsi groups.

19) *Comprehensibility of Native and Non-native German Speech, Mary Grantham O'Brien*

The only study that was found to focus on speakers of L2 German (all other studies focused of L2 English), the researchers looked at how L2 speakers of German rate other L2 German speakers and L1 German speakers. The speakers were from a university German class and

recorded and extemporaneous narration based on a picture study, that was submitted to evaluation by L2 German listeners. The rating tasks focused on phonological, fluency and linguistic resource evaluations, and found that the listeners could distinguish L2 form L1 speakers with regards to all three linguistic features. Listeners rated slower speech as less comprehensible and more heavily accented, as was the case with speech containing many pauses and phonetic errors. However, it was also found that exposure to German L2

20) *2015 Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Generalization study, Kazuya Saito, Pavel Trofimovich, Talia Isaacs*

Another study that focuses on finding what features of language are related to comprehensibility and accentedness ratings of L2 speech. The study uses L2 speakers of English who recorded extemporaneous narrations, that were then rated using 11 variables of language. Half of the listeners were L1 English expert raters and half were L1 English novice raters. The speech fragments were normalized, and transcribed so that they could be rated for accentedness and comprehensibility, as well as for several features, as was mentioned, with included pronunciation and fluency, which were rated based on the recordings, and for lexis and grammar, which were rated based on the transcriptions. Both rater groups were found to be consistent in their ratings of all linguistic features except story cohesion, and raters with linguistic and pedagogical experience compared with inexperienced raters overall. The results also demonstrate that rater experience impacts L2 speech judgements that could bias ratings in unwanted ways (experienced listeners were more lenient with comprehensibility and accentedness for example than inexperienced listeners). Accentedness ratings was found to be linked to phonological aspects of speech as opposed to comprehensibility which encompass a wider array of features including pronunciation, lexical features, grammar and discourse structure.

21) *2015 Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels, Kazuya Saito, Pavel Trofimovich, Talia Isaacs*

L1 Japanese English speakers at three levels of proficiency (beginner, intermediate and advanced), were looked at to see which speech measures listeners equated to accentedness and comprehensibility for each group. Results suggested that accent was related mainly to pronunciation features (as was shown in many of these studies over and over) and comprehensibility covered all the measures examined in the study (segmental, prosodic, temporal, lexical, grammatical) indicating a wide array of features necessary to rate comprehensibility. However, per groups, for comprehensibility for beginner and intermediate speakers, listeners were focused whether speaker attained a minimum level of segmental accuracy, fluency level, grammatical accuracy and lexical appropriateness. For intermediate and advanced speakers, listeners focused on segmental precision and grammatical accuracy. Fro accentedness for beginners and intermediate speakers, listeners focused on lexicagrammar and for intermediate and advanced speakers, listeners focused on grammatical complexity.

Segmentals, word stress, intonation and speech rates were found to be important among all three groups in terms of accentedness.

22) *2015 Lexical Profiles of comprehensibility second language speech: the role of appropriateness, abstractness and Sense relations, Kazuya Saito, Stuart Webb, Pavel Trofimovich, Talia Isaacs*

The study focuses on multiple L2 English speakers with L1 French, that were evaluated by listeners of L1 English. Speakers were in groups of beginner, intermediate and advanced language proficiency. In terms of lexis, for beginner to intermediate speakers, vocabulary (fluent, and accurate use of concrete words) was related to comprehensibility, and for intermediate to advanced speakers, comprehensibility was associated with sophisticated uses of L2 lexis (morphologically accurate use of complex, less familiar, polysemous words). The study suggests that multiple traits of lexis need to be taught for speakers in the class room in order to improve comprehensibility.

23) *2015 Lexical correlates of comprehensibility versus accentedness, in second language speech, Kazuya Saito, Straut Webb, Pavel Trofimovich, Talia Isaacs*

This study was a continuation of above study, and further found that lexical properties of speech were associated with successful L2 communication especially in terms of lexical accuracy and complexity, and for accentedness, surface-level details of lexical content (abstractness) and form (variation, morphological accuracy) were linked to accent rather than contextual details. The same measures of speech were analysed (in the form a transcript) as in above study. For details on speakers, listeners and speaking tasks, see above.

24) *2015 Second Language speakers at University: Longitudinal Development and rater Behaviour, Sara Kennedy, Jennifer A Foote and Larissa Kurtz dos Santos Buss*

The goal of the study was to evaluate what is important for non-expert raters in terms of evaluating L2 speech, and the researchers look at whether the L2 speech of university students in L2 settings develops without instruction in speaking or listening, and also how these students' speech in terms of its accentedness, comprehensibility and fluency is evaluated by community members not trained to rate L2 speech (non-expert). Overall the results demonstrate that ratings increased form year 1 to 3 except for lowest rated individuals in the first place (so there was net improvement). The study also showed that listeners themselves believed that accent could be distinct from comprehensibility as was demonstrated by this quote from one of the listeners: 'he has a think accent but he`s not hard to understand (low-rated) and she has a very thick accent but it wasn`t severe enough that it impacted my understanding of her (high-rated)'. In general, year 3 excerpts were rated significantly more favourably than Year 1, and mostly focused on accentedness and less so on comprehensibility. Ratings of segmentals were most linked to accentedness and comprehensibility and pauses and rhythm were more linked to fluency.

25) *2015 The assessment of foreign accent and its communicative effects b naïve native judges vs. experienced non-native judges, Francisco Gallardo del Puerto, Maia Luisa Garcia Lecumberri, Esther Gomez Lababex*

Because previous studies have been sceptical of using L2 listeners as reliable judges of L2 accentedness and comprehensibility, the researchers in this study compared expert L2 listeners with novice L1 listeners. Speakers were high school students from Spain studying English as a foreign language. They recorded extemporaneous narratives passed on pictures from Mayer's (1969) *Frog, where are you?.* Listener were expert L2 speakers who were teachers with extensive training, and novice listeners who were L1 speakers of English from Britain with no background in English teaching. In this instance, expert and non-expert listeners were able to evaluate speakers very similarly, but non-native judges in fact had an easier time understanding speakers than Native judges.

26) *2016 The relative contribution of segments and intonation to the perception of foreign-accented speech, Joan Sereno, Lynne Lammers, Allard Jongman*

This study looks at the relative impact of segmentals and intonation on accentedness, comprehensibility and intelligibility (specifically of L1 Korean accented English speakers), and for the task 2 Korean and 2 English speakers recorded 40 English sentences that were manipulated by combining segmentals from 1 speaker with intonation from another creating for versions: one English control, one Korean control, one English segmental with Korean intonation, and one Korean segmental with English intonation. 40 L1 English speakers transcribed and then rated for accent and comprehensibility. Finally, the results show that segmentals had a significant effect on accentedness, comprehensibility and intelligibility but intonation only had an effect on intelligibility, and this study separates segments from intonation because possibly according to the study, segmental information contributes more to the perception of foreign accentedness than intonation (based on listener ratings).

References

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561-613. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85502729?accountid=10246

Ballard, L. (2013). *Student attitudes toward accentedness of native and non-native speaking English teachers* (Master's Thesis, Order No. AAI1536504). Available from Linguistics and Language Behavior Abstracts (LLBA). (1530415508; 201405375). Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/1530415508?accountid=10246

Burda, A. N. (2000). *Language and age variables affecting measures of intelligibility, comprehensibility, and accentedness* (Doctoral Dissertation). Available from Linguistics and Language Behavior Abstracts (LLBA). (85528189; 200110737). Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85528189?accountid=10246

Chuang, H. (2010). *Undergraduates' perceptions and attitudes toward the foreign accentedness of international teaching assistants in the U.S.A* (Master's Thesis, Order No. AAI1477394). Available from Linguistics and Language Behavior Abstracts (LLBA). (852896017; 201101056). Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/852896017?accountid=10246

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal, 99*(1), 80-95. doi:10.1111/modl.12185

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly, 49*(4), 814-837. doi:10.1002/tesq.203

Del Puerto, F. G., Lecumberri, M. G., & Lacabex, E. G. (2015). The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics (United Kingdom), 25*(2), 202-224. doi:10.111/ijal.12063

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85647074?accountid=10246

Derwing, T. M., & Munro, M. J. (2009a). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 66*(2), 181-202. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/744444909?accountid=10246

Derwing, T. M., & Munro, M. J. (2009b). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476-490.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study.

Language Learning, 64(3), 526-548. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/1650526256?accountid=10246

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*(4), 655-679. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85609333?accountid=10246

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustic Society of America, 97*(5), 3123–3134.

Genesee, F. (1978). A longitudinal evaluation of an early immersion school program. *Canadian Journal of Education/Revue Canadienne de l'Éducation*, 31–50. doi:10.2307/1494684

Hsieh, C. (2011). *Rater effects in ITA testing: ESL teachers' versus American undergraduates' judgments of accentedness, comprehensibility, and oral proficiency* (Doctoral Dissertation, Order No. 3450935). Available from ProQuest Dissertations & Theses. (865328307). Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/865328307?accountid=10246

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*(2), 135-159. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/1541992732?accountid=10246

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475-505.

Isaacs, T., Trofimovich, P., & Foote, J. (submitted). Developing a user-oriented second language comprehensibility scale for english-medium universities.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*(2), 301-315. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/753819723?accountid=10246

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*(4), 554-566. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/856397647?accountid=10246

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes,64*(3), 459-489. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85661439?accountid=10246

Kennedy, S., & Trofimovich, P. (2013). First- and final-semester non-native students in an English-medium university: Judgments of their speech by university peers. *Language Learning in Higher Education, 3*(2), 283–303.

Kennedy, S., Foote, J., & dos Santos Buss, L. (2015). Second language speakers at university: Longitudinal development and rater behaviour. *TESOL Quarterly, 49*(1), 199-209. doi:10.1002/tesq.212

Kim, T. (2008). Accentedness, comprehensibility, intelligibility, and interpretability of NNESTs. *CATESOL Journal, 20*(1), 7-26. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/919961986?accountid=10246

Kramsch, C. (2009). *The multilingual subject*. Oxford.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *Journal of Multilingual and Multicultural Development.* doi:10.1080/01434632.2013.767340

Kubota, R. (2016). The multi/plural turn, postcolonial theory, and neoliberal multiculturalism: Complicities and implications for applied linguistics. *Applied Linguistics, 37(4).* doi:10.1093/applin/amu045

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Munro, M. J. & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(Suppl 1), 285–310.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, *20*, 73–97.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73-97. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/58226988?accountid=10246

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language & Speech, 38*(3), 289-306.

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*(4), 451-468. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/85532093?accountid=10246

Munro, M. J., Derwing T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*(1), 111-131.

Nguyen, B. B. (1993). Accent discrimination and the test of spoken English: A call for an objective assessment of the comprehensibility of non-native speakers. *California La Review, 81*(5), 1325.

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and non-native German speech. *Language Learning, 64*(4), 715-748. Retrieved from http://0-search.proquest.com.mercury.concordia.ca/docview/1660012159?accountid=10246

Saito, K, Trofimovich, P. & Isaacs, T, (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*(4), 439-462.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*(2), 217-240. doi:10.1017/S0142716414000502

Saito, K., Webb S., Trofimovich P., & Isaacs, T. (2016b). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition, 19*(3). doi:10.1017/S1366728915000255

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016a). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness and sense relations. *Studies in Second Language Acquisition, 38*(4), 677-701. doi:10.1017/S0272263115000297

Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics, 37*(2), 303-322. doi:10.1017/S0142716414000575

Southwood, M. H., & Flege, J., E. (1999). Scaling foreign accent: Direct magnitude estimation versus internal scaling. *Clinical Linguistics & Phonetics, 13*(5), 335-349.

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, 28I*(1), 1-30.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905-916.

Trofimovich, P., Isaacs, T., Kennedy, S., & Saito, K. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*(1), 122–140.