

Winter 2-27-2018

# College Student Rankings of Multiple Speakers in a Public Speaking Context: a Language Attitudes Study on Japanese-accented English with a World Englishes Perspective

John James Ahlbrecht  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Anthropological Linguistics and Sociolinguistics Commons](#), and the [Applied Linguistics Commons](#)

**Let us know how access to this document benefits you.**

---

## Recommended Citation

Ahlbrecht, John James, "College Student Rankings of Multiple Speakers in a Public Speaking Context: a Language Attitudes Study on Japanese-accented English with a World Englishes Perspective" (2018). *Dissertations and Theses*. Paper 4334.  
<https://doi.org/10.15760/etd.6227>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

College Student Rankings of Multiple Speakers in a Public Speaking Context:  
A Language Attitudes Study on Japanese-accented English with a World Englishes  
Perspective

by

John James Ahlbrecht

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Arts  
in  
Teaching English to Speakers of Other Languages

Thesis Committee:  
Kimberley Brown, Chair  
Jennifer Mittelstaedt  
Keith Walters

Portland State University  
2018

## Abstract

This language attitudes study used a matched guise technique to compare participant reactions of American-accented English to Japanese-accented English. Participants ( $n = 40$ ) were college educated adults living in the Portland area who completed an online survey which measured characteristics related to Status, Solidarity, and Dynamism using semantic differential Likert scales. Results showed that while Japanese-accented English received less favorable ratings on the Status and Solidarity dimensions on a statistically significant level, the small effect size may have indicated that the differences were negligible. Interpreting the results from the data through the World Englishes Kachruvian paradigm, it is argued that English learners and users would benefit by focusing more on achieving intelligibility than on attaining perfect control of an idealized variety of English.

## Dedication

I dedicate this thesis to my adviser Dr. Kimberley Brown. I would not have been able to finish this study without the countless times that she helped me along the way. I am forever in her debt.

## Table of Contents

Abstract .....	i
Dedication .....	ii
List of Tables .....	iv
List of Figures .....	v
Introduction.....	1
Literature Review.....	2
Methodology.....	25
Results.....	39
Discussion.....	46
References.....	62

## List of Tables

Table 1 – 2x2 Research Design .....	26
Table 2 – Dimensions and Scales .....	36
Table 3 – Means Scores and Standard Deviations.....	40
Table 4 – Speaker Origin Survey Question – American English .....	43
Table 5 – Speaker Origin Survey Question – Japanese English.....	43

## List of Figures

Figure 1 – Ratings by Dimension and Language Condition.....	40
Figure 2 – Status Dimension Rating by Individual Scale .....	44
Figure 3 – Solidarity Dimension Rating by Individual Scale .....	44
Figure 4 – Dynamism Dimension Ratings by Individual Scale.....	45

## Introduction

This thesis examines how a sample of proficient English users react to and evaluate two varieties of English: American-accented English and Japanese-accented English. My motivation to do a language attitudes research study on Japanese-accented English came to me in three parts. The first was shaped from my experiences with being a language learner and eventually forming a Japanese-speaking American identity. The second came from intercultural studies, sociolinguistics, and World Englishes, topics that were of particular interest to me during my participation in the MA TESOL graduate program at PSU. The third source manifested as a desire to combine the first two in the form of research that improves our understanding of how Japanese-accented English is perceived and then to apply what I learned in the process towards English teaching. It is my desire as a language learner, an English teacher, and researcher to add more data and findings to a small handful of language attitude studies that have looked closely at evaluations of Japanese-accented English.



## Literature Review

This literature review is divided into three sections. In the first section I discuss and define how I use the terms Standard Language Ideology, dialect, and accent in order to situate the independent variable of the study – English pronunciation – as an important topic of linguistics to conduct research about language attitudes. The second section provides relevant information on Kachru’s (1985) World Englishes paradigm as a strategy of avoiding ethnocentric biases in sociolinguistics research, specifically as when it pertains to language attitude studies. In the final section I first discuss what language attitude studies are and how they are conducted generally. I then describe studies similar to the one that I conducted while pointing out problems in the studies that I attempted to avoid in my own research. It is hoped that these three pieces are what is necessary and sufficient to carry out a language attitude study with a World Englishes perspective that adds useful data to the fields of sociolinguistics and social psychology.

### Standard Language Ideology, Dialects, and Accent

In this section I discuss why “Standard Language,” dialect, and accent present a challenge for discussions in general as well as in linguistic literature. I provide this discussion first as a space to operationalize the terms according to their linguistic definitions and second as an argument in favor of the World Englishes paradigm, which I describe in the next section.

Standard language ideology.

Matthews (1997) defines a “Standard Language” as a form or variety “which is learned and accepted as correct across a community or set of communities in which others are also used: e.g. Standard English, as used especially in writing, vs regional

dialects, creoles based on English, etc” (p. 352). A “Standard Language” is the result of “standardization,” which is a “process, often in part at least deliberate, by which standard forms of a language are established. Forms and varieties which are not standard are simply ‘non-standard’” (p. 352). The definition, at least, of a “Standard Language” would seem to be the same in both its popular usage and its linguistic usage. The difference, however, is that there seems to be a tendency in its popular usage to accept “Standard Language” as fact, whereas the linguistics standpoint is to consider a “Standard” form as an ideology held in place by the speakers of a language.

Problems arise, however, when the term “non-standard” is used to describe any forms that differ from the “standard.” “Non-standard” might be interpreted as “sub-standard,” suggesting a difference in quality. Assuming the language works for the people who use it, however, there simply is no difference in the quality of a language. Edwards (2009) attempted to explain this argument with a discussion on cultural relativism:

I take it for granted that a society which condones female circumcision, believes in witchcraft, and eats its enemies is inferior, in these respects at least, to one which does not. I do not see that this constrains me to accept, as well, that the language of that society is inferior to the one spoken next door, even if the neighbors are all feminists, scientists, and vegetarians. (p. 51)

The point is made here to acknowledge the prison house of language, where sometimes the term “non-standard” is the most accessible term to use when referring to linguistic forms which diverge from some norm, whatever that may be. In this paper I have tried to avoid using the term “non-standard,” using instead World Englishes terms (discussed below). However, I use the term in some places where other researchers frame their

discussion around a “non-standard” form and I use quotation marks to indicate that I am using the term as an ideology, rather than as a fact.

Dialect.

When linguistic variation occurs due to regional differences, the term usually used is *dialect*. The popular usage of *dialect*, however, has come to have a negative connotation. “In some speakers’ minds, to say that people speak a dialect is tantamount to saying that they are provincial, perhaps not well educated – though this is neither a necessary nor a proper connotation of dialect in its technical meaning” (Kachru & Nelson, 2001, p. 10). An argument that is often used by linguists to point out the difference between a language and a dialect was amended by Lippi-Green (2012), who wrote that “Max Weinreich is widely quoted as pointing out that a language is a dialect with an army and a navy; I would like to add to that observation that a dialect is perhaps nothing more than a language that gets no respect” (p. 46).

The technical definition of *dialect* is “any distinct variety of a language, especially one spoken in a specific part of a country or other geographical area. The criterion for distinguishing ‘dialects’ from ‘languages’ is taken, in principle, to be that of mutual intelligibility (Matthews, 1997, p. 96). The issue of mutual intelligibility is important here in its connection with the linguistic term *dialect continuum*, which is “a progressive shift from one form of speech to another across a territory, such that adjacent varieties are mutually intelligible, but those at the extremes are not” (p. 96). In the case of English, as a language spoken around the world, what happens when territories where the same language is used are not adjacent, but in distant areas around the world? One does not often hear of an “American dialect of English” unless the language in question is

referring to a “non-standard” form. Likewise, one rarely hears about a “Chinese dialect of English,” but this is likely because a Chinese person speaking English is often said to have a “Chinese accent.”

I offer the definitions of *dialect* and *dialect continuum* here to illustrate their relationship with English as a global language. On the one hand, *dialect* would be well-suited to describe varieties of English used in various parts of the world where it is assumed that there is at least some mutual intelligibility between them, since they are all the same language. On the other hand, Matthews’ definition of *dialect continuum* assumes that distance is a factor in determining mutual intelligibility. While this definition is certainly true of English in some contexts, there are many cases in which distance between territories factors in very differently. A person from Saudi Arabia could communicate with someone from South Korea if they were both proficient users of English, for example. That the two countries are not neighbors or “adjacent” would be irrelevant. If it were not for the distance component of *dialect*, it would be a useful term to operationalize the varieties examined in this paper (i.e., English spoken in America and Japan). Due to the discrepancy between the definitions, and because *dialect* does not seem to be applied to Englishes spoken around the world, however, the term appears only minimally throughout this paper.

Accent.

Just as there is a difference in usage between the popular and linguistic usages of “Standard Language” and *dialect*, there is also a disparity in the uses of the term *accent*. In the case of *accent*, however, the difference to note for this paper is not in determining if it is an ideology or fact, or in the meaning of the word, but in who it applies to:

As feminist theorists have pointed out, everyone has a gender, but the hidden norm in law is male. As critical race theorists have pointed out, everyone has a race, but the hidden norm in law is white. In any dyadic relationship, the two ends are equidistant from each other. If the parties are equal in power, we see them as equally different from each other. When the parties are in a relationship of domination and subordination we tend to say that the dominant is normal, and the subordinate is different from normal. And so it is with accent...People in power are perceived as speaking normal, unaccented English. Any speech that is different from that constructed norm is called an accent. (Matsuda, 1991, p. 805)

In the Linguistics Dictionary, *accent* is given the following definition:

A variety of speech differing phonetically from other varieties; thus, as in ordinary usage, 'a Southern accent,' 'Scottish accents.' Normally restricted by linguists to cases where the differences are at most in phonology: further differences, e.g. in syntax, are said to be between dialects. (Matthews, 1997, p. 4)

As I use *accent* frequently throughout this paper, I use the term as defined by Matthews only as a means to refer to an identifiable set of phonological features. For example, I attach *accent* to all varieties that I refer to in this paper. English spoken by Americans is operationalized as American-accented English and English spoke by Japanese people is operationalized as Japanese-accented English. In other words, America does not receive any special privileges which could be inferred from being called "unaccented." I discuss this further in the following World Englishes section.

### World Englishes

Crystal (2003) claimed over ten years ago that ESL, EFL, and EIL (English as a second, foreign, and international language, respectively) speakers easily outnumbered ENL (English as a native language) speakers (pp. 61, 69). "Accepting even cautious estimates, there must be at least three non-native users of English for every old-country native user" (Kachru & Nelson, 2001, p. 14). Assuming the trend has not changed over the last 10 to 15 years, now more than ever there is an urgent necessity for any English

users who are in contexts where they come into contact with different cultures to become accustomed to hearing different kinds of English.

The necessity for English users to familiarize themselves with different varieties of English begins from the false assumption that every English user is going to sound like an American or British speaker. To put it another way, the dichotomous way of thinking, where there are “native speakers” and “nonnative speakers,” has become problematic for English users. Not only is an exact definition of “native speaker” nearly impossible to pin down (Paikeday, 1985), it has been pointed out that “it is also inappropriate to use the label *non-native speaker* for a person who has learned English as a second or foreign language and achieved bilingual status as a fluent, proficient user” (Tokumoto & Shibata, 2011, p. 393). Reaching agreement on exactly what it means to be a “fluent, proficient user” of a language is also likely challenging for any two speakers. However, individuals who grew up speaking Spanish at home and English everywhere else, for example, would be just as fluent and proficient in English as another person who grew up using English both in public and at home. Likewise, someone in an English as a foreign language context (i.e., Japan), might also be able to attain a similar level of fluency through practice and determination. The difference between a native, second, and foreign language may be easier to define in some linguistic contexts, but it might also be argued that the situation with English as a global language may require a refining of these terms.

While the Kachruvian World Englishes paradigm is still a marked viewpoint with regard to the English language in general, in English language teaching and even within the field of linguistics, it has arguably gone the furthest in exploring and explaining what is happening with English globally. Kachru (1985) posited that the universalization,

internationalization, stratification, and “spread of English may be viewed in terms of three concentric circles representing the types of spread, the patterns of acquisition and the functional domains in which English is used across cultures and languages” (p. 12). Kachru labeled these circles the Inner Circle, which refers to the traditional bases of English – the regions where it is the language of the majority (e.g., the USA, the UK, and Australia); the Outer Circle, which includes regions where (a) English is only one of two or more codes in the linguistic repertoire of such bilinguals or multilinguals, and (b) English has acquired an important status in the language policies of most of such multilingual nations (e.g., India, Nigeria, and Singapore); and the Expanding Circle, where English is recognized as an important international language (e.g., China, Japan, and Indonesia) (pp. 12-13). The Linguistics Dictionary defines an international language as one which is “used internationally; specifically, one recognized officially as a language of international meetings, diplomacy, etc” (Matthews, 1997, p. 183). While Matthews’ definition is useful for many English-speaking contexts, it can also be expanded to apply to day-to-day interactions between speakers, such as small business transactions at a grocery store, for example.

There is often a tendency to think of the Inner, Outer, and Expanding circles as corresponding to regions where English is regarded as a native, second, and foreign language, respectively. This may be an oversimplification, however:

The concept of English in its inner, outer, and expanding circles is only superficially equivalent to *native*, *ESL*, and *EFL*. In thinking of a country as an ESL country or of a person as an ESL speaker, for example, we perpetuate the dichotomy of native versus non-native, “us versus them.” (Kachru & Nelson, 2001, p. 14)

A World Englishes perspective proposes a way to think of English that does not create Others of everyone outside of places where English is dominant. An Inner Circle variety, for example, is not necessarily the target of English speakers/learners in either Outer or Expanding circles. In the Outer Circle, bilinguals/multilinguals proficient in English, who rarely/never interact with speakers from the Inner Circle, are so common that the English language has become part of their identity, just as it is for Inner Circle speakers. It is not a safe assumption at all that Outer Circle English users wish to speak an American or British variety. Finally, in the Expanding Circle, where English is often referred to as a foreign language, English-user identity is sometimes either tenuous (Matsuda, 2003) and an Inner Circle variety is often the target, or the ability to speak English is essentially reserved to an elite group. In either context, however, it may be more accurate to define English as an International language.

It has been argued that categorizing people into native speakers and non-native speakers “is problematic in terms of equality and power relations among English users, whereby those who speak English as their native language are automatically in a position of power as compared with those who have to learn it as a second or foreign language” (Tokumoto & Shibata, 2011, p. 392). This may just be a social fact in most sociolinguistic contexts, but from a global perspective and English’s status as the world’s first truly global language (Crystal, 2003), it may be prudent for “native speakers” to not assume that they “are automatically in a position of power.” The World Englishes ideology proposes that “English now belongs to all who use it” (Kachru, 1985). What Kachru’s ideology suggests is that Inner Circle speakers can still take ownership of English, but they might also benefit from avoiding the attitude that their English is



“better” or more “correct” when speaking with Outer and Expanding Circle speakers. Outer and Expanding Circle speakers also benefit from the World Englishes ideology by focusing less on emulating Inner Circle speakers’ pronunciation and more on intelligibility. In short, the ideology proposes that all speakers of English put less effort into deciding who is speaking the best and more energy into communicating effectively. However, owing to language attitude studies, it can be said with relatively high certainty that many Inner, Outer, and Expanding Circle users of English do not subscribe to a World Englishes ideology.

Although many language attitude studies are not conducted within the World Englishes framework, research on language attitudes has gone a long way in informing us about how people evaluate dialects, accents, and/or varieties of a language. In the interest of staying within the World Englishes paradigm where possible, rather than relying exclusively on accent, standard/non-standard English, and native/nonnative speaker, Inner, Outer, and Expanding circles are used in this paper, where applicable, to refer to corresponding varieties/regions where the terms are appropriate.

### Language Attitudes

Language attitudes is a field of study within sociolinguistics and social psychology that became prominent over 50 years ago, due initially from research conducted by Lambert (Lambert, Hodgson, Gardner, & Fillenbaum, 1960). Before reviewing language attitude studies more generally, the term *attitude* is discussed below with the intention of clarifying its definition within this field.

The mentalist view on attitudes is what language attitude research is usually based on (Garrett, 2010). Social psychologists in this view often divide attitudes into three

components. In terms of Allport's definition of attitudes as "a learned disposition to think, feel and behave toward a person (or object) in a particular way" (as cited in Garrett, 2010, p. 19), attitudes have a cognitive, affective, and behavioral component (Fasold, 1984; Garrett, 2010; Walters, 2007). "Cognitive responses involve beliefs, and affective responses involve emotions, feelings, and sympathetic nervous system activity, while behavioral responses involve overt actions" (Walters, 2007, p. 650).

Each component of attitudes at least partially indicates what kind of attitude a person might have, but it cannot always be assumed that what a person thinks about a language variety reflects or predicts how they feel or how they will act when they hear it. For example, an Inner Circle speaker might *say* that they have no issue with "foreign" accented English, but then she/he quickly gets frustrated when speaking with a foreign-accented person, and may even get angry, specifically about the accent, when communication feels difficult. The reverse is also a possibility, such as someone claiming that they strongly dislike foreign accents but they are neither noticeably affected nor behave negatively towards someone with a foreign accent.

When cognition and affect are in disagreement, it is difficult to predict which one a person will act on. More often than not, it may be the context that determines the outcome. All else being equal, however, "whether the cognitive or affective response wins out often influences behaviors, for example in assessing whether a speaker of [a] stigmatized variety is employable" (Walters, 2007, p. 650). This illustrates how a decision between thoughts and feelings can have life-changing consequences in a context of language attitudes.

Researchers have employed various methods of investigating language attitudes. The following section discusses how research is conducted to investigate conscious attitudes (the cognitive component) and unconscious attitudes (the affective component). Although the behavioral component of language attitudes has been investigated in societal treatment studies (Garrett, 2010, p. 142), by observation (Fasold, 1984, p. 152), and discriminatory linguistic profiling that occurs in courtrooms (Baugh, 2003; Purnell, Idsardi, & Baugh, 1999), an in-depth discussion of this type of research is beyond the scope of this study. Focus is mainly on the method of investigating attitudes at the unconscious level since the current study used the Matched Guise Technique, “an experimental method that has become virtually standard in language research, in either original or modified form” (Fasold, 1984, pp. 149-150).

It should be noted that Fasold made this claim over 30 years ago and that some scholars have advised caution in interpreting the results of the quantitative data from these studies. Walters (2007) suggested three important matters to look out for in matched guise studies. First was that the sample populations often are not representative of a more general population. Second was that researchers frequently do not have the expertise in questionnaire-based studies, psychology, and statistics to make sound generalizations from the data. Third was how “the research cited and models used data from several decades ago, and researchers have generally not kept abreast of empirical and theoretical work in sociolinguistics or social psychology and psychology more broadly, whether work on the nature of attitudes or the complexity of constructing valid surveys” (pp. 652-653).

On the other hand, researchers have been conducting many language attitude studies since Fasold claimed it as a standard practice, including within the last 10 years (e.g., Bauman, 2013; Chien, 2014; Garrett, 2010; Speelman et al., 2013; Tokumoto & Shibata, 2011), and it is a bona fide technique that may still have something to tell us. However, just as this study itself questions the label of “standard,” it is advisable to be a critical consumer of language attitude research due to the difficulty in controlling all of the variables sufficiently.

Two types of language attitude studies.

Although there is considerable methodological variation in language attitude studies, the two most frequently used approaches are *direct measures* and *indirect measures*. As the current study uses indirect measures, the direct approach is described only in how it compares with the indirect approach while the indirect approach is discussed in detail.

The direct method.

The direct method can be loosely associated with the cognitive component of attitudes. “A totally direct method would require subjects to respond to a questionnaire or interview questions that simply ask their opinions about one or another language” (Fasold, 1984, p. 149). These kinds of studies “typically measure consciously and deliberately constructed and expressed attitudes” (Speelman, Spruyt, Impe, & Geeraerts, 2013, p. 3). According to Garrett (2010), “the direct approach has probably been the most dominant paradigm if one looks across the broader spectrum of language attitudes research” (p. 159). Walters (2007), on the other hand, pointed out that “many find questionnaire-based studies suspect, contending that their findings are best taken as

evidence of overt or imagined norms rather than actual behaviors” (p. 651). In other words, meta-linguistic questions yield meta-linguistic answers, and when someone is answering questions on this level their thought processes might be markedly different from what happens in everyday interactions. As Walters (2007) explains, language users are often “unable (and sometimes unwilling) to report accurately their own attitudes and behaviors or those of others” (p. 652), and their responses are “susceptible to social desirability or self-flattering strategies” (Speelman, Spruyt, Impe, & Geeraerts, 2013, p. 3). This does not mean that findings from direct approach studies are unimportant. “Rather, such reports are taken as evidence of the robust structuring power of language ideologies” (Walters, 2007, p. 652). This kind of research is an effective method for revealing what people are thinking about languages at a conscious level. In order to gain access to the unconscious level and investigate how people *feel* about different language varieties, however, the indirect method is often used.

The indirect method.

As the direct method correlates to the cognitive component of attitudes, the indirect method is often an investigation into affect or the subconscious. “A totally indirect method would be designed to keep the subject from knowing that her language attitudes were being investigated” (Fasold, 1984, p. 149). The matched guise technique (MGT), originally developed by Lambert (1960), is one of the most frequently used methods of indirectly measuring language attitudes. The MGT was an innovative method of investigating attitudes related to language:

The MGT involves the presentation of various audio fragments that are recorded in different language varieties by one speaker. The main idea is that the recorded accents unfold impressions of personality traits (which are, in fact, impressions of

different language varieties since all recordings are made by the same speaker) to listeners who are supposed to be unaware of the fact that only one speaker is involved. (Speelman, Spruyt, Impe, & Geeraerts, 2013, p. 4)

Fasold (1984) further explains the methodology:

To achieve this, a number of bilingual [or bi-dialectal] speakers fluent in the languages [or dialects] under investigation are recruited. These speakers are tape-recorded reading exactly the same passage, once in one language [or dialect] and once in the other. The passages are arranged on a tape-recording in such a way that it appears that each passage has been recorded by a different speaker. (p. 150)

For example, if a study is investigating responses to American English and Indian English, then one speaker who can use both varieties is recorded once in both varieties. When participants listen to the recordings, however, the first recording may be of the American English guise, but the second and third recordings are different speakers using either the same varieties (e.g., American or Indian English) or an entirely different variety (e.g., Singapore English). For the fourth recording the listeners hear the first speaker again, but this time in the Indian English guise. “By this time, listeners will presumably have forgotten the voice quality of the first speaker and will take it that the fourth speaker is someone they haven’t heard before” (Fasold, 1984, p. 150).

Although the MGT has been called an indirect method of investigating language attitudes, it may be important to point out that it is not, strictly speaking, completely indirect. “The matched guise technique is direct in the sense that the listeners are explicitly asked to give their opinions of the speaker’s characteristics. It is indirect in the sense that listeners are asked to react to *speakers*, not languages, and they are not aware that they are hearing the same person in each guise” (Fasold, 1984, p. 150).

It is also possible to use the MGT in a more direct way. For instance, asking respondents to answer open-ended questions that are specifically about the speaker’s

language or accent would be more direct and conceivably more about investigating cognitive attitudes than affective attitudes. This discussion is beyond the scope of the current study, however.

While the current study used a MGT, the *Verbal-Guise Technique* (VGT), a close variant of the MGT, should be briefly discussed here because of its similarity to the MGT and because studies which used the VGT are discussed in later sections.

A VGT is where recordings are made with multiple speakers. One of the most important differences between an MGT and a VGT is that speakers in VGT studies make only one recording and use the language variety that they speak the most naturally (Chiba, Matsuura, & Yamamoto, 1995; Chien, 2014; Hosoda, Stone-Romero, & Walter, 2007; Lindemann, 2003; Podberesky, Deluty, Feldstein, 1990; Smith & Bisazza, 1982; Smith & Rafiqzad, 1979). The MGT can be said to be artificial in that it asks a speaker to use a language variety that they likely do not use every day, while the VGT asks speakers to speak only in the variety that they are the most comfortable with. On the other hand, the VGT also increases the variables, such as different voice quality, accent strength, rate of speech, etc. The difficulty in controlling these variables was the reason for using an MGT for the current study.

Another common variation to matched and verbal guise studies is to make the recordings sound more like spoken speech by having speakers talk about a specific subject, such as a description of the speaker's home (e.g., Bauman, 2013). For tighter control of variables, however, the traditional method created by Lambert (1960) was to have each recording be the same content read by each speaker. This issue is further discussed in the methodology section.

The three evaluative dimensions.

Participants in language attitude studies are typically asked to listen to and evaluate recordings on a series of semantic differentials (Osgood, Suci, & Tannenbaum, 1969) presented as Likert scale items. The items are often further organized into different dimensions, although the dimensions are not always labeled the same. Zahn and Hopper (1985), using a factor analysis, found the 30 most salient terms that were frequently used in language attitude studies and divided them into three dimensions: superiority, attractiveness, and dynamism (pp. 117-118). As superiority and attractiveness are descriptively loaded terms, these dimensions are often labeled as Status and Solidarity, respectively. Status is defined “as a blend of social status, intellectual achievement, and the speech characteristics of advantaged and educated members of society” (Zahn & Hopper, 1985, p. 119). Solidarity is operationalized as “social attractiveness, solidarity, trustworthiness, character, benevolence, likeability, and aesthetic quality” (p. 119). The third dimension, Dynamism, which does not appear in nearly as many studies, is used to measure “speakers’ social power, activity level, and the self-presentational aspects of speech” (p. 119).

Related research.

I searched for language attitude studies that looked specifically at Japanese-accented English and was only able to find one study (Cargile & Giles, 1998) which I will discuss later in this section. However, there have been an abundance of MGT and VGT studies which compared rater evaluations between Inner Circle varieties of English with Outer or Expanding Circle “Asian” Englishes. The operationalizing of “Asian English” is problematic, to say the least, in that it puts several linguistically diverse



groups of people together. Despite the fact that “the eight most populous Asian nations are home to about 3.5 billion persons, the U.S. government has only one term that lumps all of these nations, ethnicities, cultures and languages from the Far East, Southeast Asia, and the Indian subcontinent together: Asian” (Lippi-Green, 2012, p. 282). The inherent problem with “Asian” as a group of people aside, studies using this term can still be useful in revealing language attitudes.

Lindemann (2003) conducted a verbal guise study in which 39 undergraduates at the University of Michigan evaluated Korean-accented and American-accented English. Lindemann described the participants as “native speakers of English” who were “mostly first-year students recruited from introductory classes in linguistics and history” (p. 352) and “reported little experience with Koreans on the language background questionnaire” (p. 357). Lindemann did not report information on the participants’ ethnicity.

Participants listened to ten recordings of a 101-word text, presented as an outgoing answering-machine message for a doctor’s office. The voices on the recordings included two male and two female Korean-accented speakers, two male and two female American-accented speakers, one female British-accented speaker, and one male “identifiably non-Midwestern” American-accented speaker (p. 352). Lindemann collected and analyzed the data from the four Korean-accented speakers and the four American-accented speakers. However, because the British-accented speaker and the “non-Midwestern” American-accented speaker were added as distractor voices, and one of these voices was always played first for the participants “in order to give them practice with the scoring procedure” (p. 353), Lindemann did not collect and analyze the data for these voices. After hearing each of the voices twice, participants rated each speaker on

six status-related and six solidarity-related characteristics, with judgements reported on seven-point Likert-scales from “very” to “not at all” (p. 353).

An additional component to Lindemann’s study was to have the participants assess the ethnicity of each of the speakers in order to ascertain how listeners’ identifications of the speakers’ ethnicity may relate to salient social groups for the listeners. This data was gathered after the initial evaluations by having participants listen to each of the recordings a third time, rate the voices on a 7-point scale “as to how likely it was that s/he was a native speaker of English” (p. 353) and then assess the speakers’ ethnicity in an open-ended format. As a final task, participants were asked to fill out a language background questionnaire which asked them to “list countries and languages of non-native speaker friends, teachers, and co-workers” (p. 353). While only 8% of the participants correctly guessed “Korean,” over 50% wrote down “Korean,” “Chinese,” “Japanese,” or “Asian.” This finding suggests that the participants either could not discern the difference between these groups or they simply grouped all Asians together, consistent with Lippi-Green’s (2012) observation “that the concept *Asian* evokes an association not with a specific nation or geographical region, but with race” (p. 285).

With respect to the evaluations, compared to American-accented English, Korean-accented English was rated lower on the status dimension but there was no significant difference on the solidarity dimension. “The evaluation of the Korean speakers as equally friendly but less intelligent than the U.S. speakers is rather different from what we might expect in view of widespread stereotypes of Asians as hardworking, intelligent, and reserved” (Lindemann, 2003, pp. 357-358). These findings indicated that the “listeners’

higher-order indexical reactions thus appear to identify a generalized ‘foreign faultiness’ rather than a relationship between specific features and speaker traits” (pp. 358-359).

Hosoda, Stone-Romero, and Walter (2007) also used a verbal guise to compare American-accented English and Vietnamese-accented English. The researchers for this study unintentionally illustrated the standard language ideology by operationalizing American-accented English as “nonaccented English” (p. 307). Additionally, they operationalized Vietnamese-accented English as “Asian-accented English” (p. 307), exemplifying Lippi-Green’s (2012) assertion about the tendency to put all Asians into one group.

Participants were 97 (70 female, 27 male) college students at a university in northern California. The researchers noted that the “sample was diverse in terms of its ethnic composition” (p. 312) with 39% Asian ( $n = 38$ ), 23% Euro-American ( $n = 22$ ), 18% Hispanic American ( $n = 17$ ), 4% African American ( $n = 4$ ), and 15% of mixed-ethnicity ( $n = 15$ ) (pp. 312-313).

The procedure involved experimental sessions in which groups of 3 to 8 participants were randomly assigned to one of four language conditions: a “nonaccented” male, a “nonaccented” female, an “Asian-accented” male, or an “Asian-accented” female. Participants listened to a one-minute recording of their speaker describing their favorite things to do, a recent movie they saw, and an experience related to a first job (p. 313). After listening to the speaker, participants filled out a 59-item, 7-point bipolar adjective scale on attributes that pertain to social status, “social competence,” and traits “considered stereotypical of Asians” (p. 314). After the 59 items, 4 additional semantic differential scales were added to measure the speaker’s accent. Finally, the participants

were also asked to “write down the perceived ethnicity of the speaker to whom they listened” (p. 315).

Similar to the Lindemann (2003) study, “89% of participants ( $n = 42$ ) in the Asian-accented English condition correctly identified the speakers as either Asian ( $n = 30$ ) or of an Asian nationality ( $n = 12$ ; 6 as Vietnamese, 2 as Chinese, 1 as Japanese, 1 as Asian Indian)” (Hosoda, Stone-Romero, & Walter, 2007, p. 315).

Hosoda et al. hypothesized that “compared to standard American English-accented speakers, listeners will evaluate Asian-accented English speakers more favorably on attributes related to social status but less favorably on attributes related to solidarity” (p. 311). The findings were nearly the opposite, however. “The Asian-accented English speakers were rated more negatively on potency and communication but more positively on concern for others and being nonthreatening” (p. 319). In this study, the “potency” factor was similar to the *dynamism* dimension described above, “communication” as part of *status*, and “concern for others and being nonthreatening” as *solidarity*.

In explaining their findings, the researchers pointed to a possible influence of their participants: 79% claimed to have friends who spoke English with an accent and 48% were either bilingual or spoke a language other than English at home (p. 320). Additionally, in respect to similar studies that had very different results, Hosoda et al. claimed that “the relatively favorable ratings of Asian-accented English speakers on the social status dimension and unfavorable ratings on the solidarity dimension might be limited to a few select Asian ethnic groups, not to all Asians” (p. 320).

In their matched guise study, Cargile and Giles (1998) contended “that Japanese-accented English is a special variety of non-standard speech in America” (p. 342). A Japanese male speaker was recorded reading from an English language script, approximately two minutes long, in four target accents: American, moderate Japanese, heavy Japanese, and heavy/disfluent Japanese. Participants were 240 non-Asian American undergraduates (222 reported as Anglo) at a major California university who were told that the speaker was a teacher’s assistant on campus and that the recording was from one of his recent mini-lectures. Cargile and Giles claimed that the findings from their data demonstrated that a moderate Japanese accent compared favorably to an American accent on status-related traits, but unfavorably on traits related to attractiveness. They claimed that “this evaluative profile is relatively unique among profiles provided by American listeners” (p. 349). Arguing this point, the researchers compared their findings from several other studies, noting that American listeners consistently rated other varieties in three ways. In one category, Americans downgraded varieties such as Spanish-accented, Appalachian-accented, and “vernacular Black English” on ratings related to status, but rated these varieties equally on solidarity traits. In a second category, varieties such as German-, Norwegian-, and Italian-accented speakers were downgraded on both status and solidarity-related traits. For a third category, the Cargile and Giles pointed out that “the only speakers who compare favorably on status-related traits and unfavorably on attractiveness-related ones are British, Malaysian, and the present, moderate Japanese-accented speaker” (p. 350). This suggested, the researchers argued, that for the American listeners who participated in this

study, Japanese-accented English shares an evaluative profile with British-accented English (RP) and Malaysian-accented English:

Put succinctly, what do British, Malaysian, and Japanese-accented speakers have in common? They appear to represent varieties of non-standard speech associated with outgroups that Americans perceive as competitive, status-equals; that is, in contrast to outgroups who are perceived to be subordinate. (Cargile & Giles, 1998, p. 350)

This study closely followed the dimensions proposed by Zahn and Hopper, and even went so far as to manipulate the dynamism dimension by having a “neutral” passage and an “aggressive” passage. The data revealed a significant increase in dynamism ratings for the guise in the moderate-accent condition.

Although all three studies appear to have examined language attitudes on the same object – a linguistic variety operationalized as an “Asian-accented English” – the findings from each study contrast with one another so remarkably that it would be difficult to make any generalizations regarding an “Asian-accented” English. However, as they each dealt with a different group – Korean, Vietnamese, and Japanese – one might speculate that American English speakers react to each of these groups differently, even if they think of them all as “Asian.”

The present study also investigated language attitudes toward Japanese-accented English. A key difference in this study from the Cargile and Giles (1998) study is that although it was also a matched guise study where a speaker performed guises and respondents heard four recordings, two of the recordings were guises of the same speaker and two recordings were distractor voices. One distractor voice was an Inner Circle speaker of American-accented English, and the other was an Outer Circle speaker of Philippines-accented English.

## The Present Study

Although the methodology of the current study was considerably different from the Cargile and Giles study, it was hypothesized that results would be similar.

Research question: Will college educated participants living in Portland evaluate the guise of a female Japanese-accented English speaker in her twenties negatively on traits related to solidarity and dynamism, but more favorably or equally on status-related traits when compared with the same participants' evaluations of the same speaker in an American-accented English condition?

## Methodology

This study used a matched guise technique to test listener evaluations of American-accented English (AE) and Japanese-accented English (JE). Listeners were college level educated adults living in the Portland area. Evaluations were measured with a five-point semantic differential on three dimensions: status, solidarity, and dynamism. Data was analyzed with SPSS after it was collected online using Qualtrics survey software (Qualtrics Labs, Inc. 2011).

### Context and Participants

Respondents ( $n = 40$ ) were college educated adults residing in the Portland metropolitan area and were recruited by snowball sample, mainly by email, word-of-mouth, and social media. Participants began the survey by clicking on a link that took them to the Qualtrics website where they were first presented with an anonymous survey consent form. The second page asked participants to provide some personal information about themselves, including confirmation of being over 18, one question each to identify their gender, level of education, and proficiency in English. The third page instructed participants to listen to each recording and to use their first impressions of the speakers to make evaluations on the traits provided. Due to the software allowing participants to replay the recordings an unlimited number of times, participants were discouraged from repeating the recordings and were asked not to listen to any recording more than five times. Participants were required to answer each question before proceeding to the next page, and they were not provided with a button to go backwards in the survey.

As previous studies have shown that evaluations can be influenced by audio sample order, this study controlled for this possibility by counterbalancing. This was



accomplished by creating two separate surveys in Qualtrics. One version had the JE condition first, two distractor voices in the second and third positions, and the AE condition fourth ( $n = 22$ ), while another version was created with the AE and JE conditions reversed ( $n = 18$ ). Respondents were split into two groups, resulting in a 2x2 research design (distractors were not factored).

	<b>Condition 1 (guise)</b>	<b>Condition 2 (distractor)</b>	<b>Condition 3 (distractor)</b>	<b>Condition 4 (guise)</b>
<b>Group 1</b> $n = 18$	Japanese English	Inner Circle American English	Outer Circle Philippines English	American English
<b>Group 2</b> $n = 22$	American English			Japanese English

Table 1 – 2x2 Research Design

#### The Speech Evaluation Instrument (SEI)

As there were several steps to create the SEI for this study, this section is divided into subsections. The technical side of creating the survey is summarized, followed by a description of the matched guise speaker, how the JE and AE conditions were operationalized, the use of distractor voices, and the choice of stimuli. The final subsection describes the SEI adapted from Zahn and Hopper and the use of a semantic differential.

##### Creating the survey.

Once the four language conditions were set, the next step was to ensure that each of the recordings had similar audio quality so that listeners would not be distracted by noticeably different sound. The three speakers were all females in their early to late twenties and were each chosen for different reasons which are further discussed below in the Matched Guise and Distractor Voices sections.

Each of the three speakers sent their recordings as an email attachment. The guise and Philippines speakers used their smart phones, which resulted in recordings with clear voices and no background noise. The American speaker, however, used a PC, headset, and microphone. The difference was immediately apparent, and she was asked to re-record twice. On the third attempt, the speaker used her smart phone, and the recording was much closer in quality to the others, so this final recording was used.

The recordings were then uploaded into the Qualtrics survey. Due to Qualtrics' audio file size limit, the recordings could not be uploaded directly; instead, the audio files were converted into videos with Microsoft Movie Maker, uploaded to Youtube, and then embedded into the survey. A picture was required to make a video, so a picture of Hammerton College was downloaded from Creative Commons and appeared the same way in each video. The photo was selected for its lack of people or identifying markers and its appearance as a building that looked like it belonged on a college campus.

Upon pilot testing the survey, it was found that it worked smoothly on PCs and laptops, but when the survey was accessed with a tablet or a smart phone, the embedded videos would not appear. Embedding an additional link made it possible to view the video by switching to the Youtube app, but if the user exited Youtube and tried to continue the survey by clicking on the original link, they would be taken back to the anonymous consent form and be forced to start over. This could be avoided by directly returning to the browser from Youtube. Due to this difficulty, however, all survey invitations included a message recommending the use of a PC or a laptop.

A final issue that came up while titling the videos was that the titles needed to be organized by number due to intentional lack of description (e.g., Public speaking -

Speaker #1). However, as mentioned above that there were two versions of the survey, the guise speaker's recordings needed to be alternated. Qualtrics software did not have a feature to swap the videos, so two identical surveys were created, with the only difference being the order of the first and last recordings. This meant that two versions of the matched guise speaker's Youtube videos had to be uploaded, with one version having the JE condition named "Public speaking - Speaker #1" and the AE condition named "Public speaking – Speaker #4", and the other version had the alternatively titled videos in opposite order from the other version of the survey. It was also necessary to create an entirely different survey in Qualtrics which was identical to the first except for the video order.

The matched guise speaker.

It was carefully considered whether it would be better to use a matched or verbal guise technique for this study. Whereas a matched guise study only compares responses to two (or more) different "guises" of one bi- or multi-dialectal speaker, a verbal guise compares participant responses to multiple speakers, each using the language variation that she/he is most comfortable using. Although a matched guise decreases threats to internal validity by reducing the number of confounding variables compared to a verbal guise, the challenge is usually finding a speaker who is capable of using two or more dialects convincingly and naturally.

Due to the availability of a bi-dialectal speaker, I decided to do a matched guise. The speaker was a female graduate student of Japanese descent in her late twenties who moved to the United States from Japan when she was 12. In an interview, she described her motivation to "speak like an American" as she was growing up due to ridicule of her

Japanese accent by American classmates. She was highly recommended for this study by faculty and cohort at her university because she was said to have “an undetectable accent.” Although it is not the main focus of this study, respondents were asked to guess where they believed each speaker was from. This question was included in response to Lindemann’s (2003) study where it was observed that “matched guise studies have very seldom addressed the question as to how accurately listeners identify the accents” and that the task “was clearly a difficult one” (p. 353). Respondents’ answers in this study are analyzed in the Results and Discussion sections.

#### American-accented English.

AE is often given different labels, such as “Standard English,” “unaccented English,” and “Standard American English.” As discussed above in terms of the Standard Language Ideology, the problem with these labels is that they do not refer “to one particular language variety, but to a ‘collectively held ideal’ of unaccented, unmarked American English that is associated with educated, middle-to-upper class, white European-descended people from the American Midwest, parts of the Northeast, and the West Coast” (Lippi-Green, 2012, p. 41). Since any variations to this ideal, such as Southern English, Black English Vernacular, or English that sounds “foreign,” evoke their own evaluations, it was important that the speech in the AE condition be as close as possible to what would generally be perceived to be “Standard American English.” This was accomplished by closely monitoring the audio sample of the AE guise for any speech patterns that might indicate some level of “non-nativeness” or dialectal variation. As part of this study investigated the extent to which participants believed that the AE guise was in fact an American-accented English, and because of the unlikely scenario that every

participant would assess the speaker's English variety as that coming from an American, the more appropriate way operationalize her speech is probably as a highly fluent user of American English. I discuss this further in the Results and Discussion sections.

#### Japanese-accented English.

The primary challenge with operationalizing Japanese-accented English is that, as part of Kachru's Expanding Circle, it is at least somewhat controversial to legitimize it by calling it a variety. As the respondents were not informed of where the speakers were from, however, the more pertinent question for the methodology was to determine if the respondents could discern that they were hearing Japanese-accented English. For this condition, I asked the guise speaker to emulate a "moderate accent" (i.e., pronouncing words with a marked Japanese accent but also intelligible and discernable as English spoken by a proficient speaker).

It should be noted that the speech produced for this condition was not a way the speaker uses English on a daily basis, and as such could not be considered her most natural speaking style. However, I asked her to use this accent because I believed that she would be able to perform this speech in a way that was sufficiently natural sounding for the purposes of this study.

#### Distractor voices.

It is important in a matched guise study for the participants to not realize that they are evaluating the same voice more than once. Even though the speaker used two distinctly different varieties of English, hearing one guise after the other would have made it easier for respondents to discern two recordings as the same person. For this reason, participants were asked to listen to and evaluate two distractor voices between the

two guises. Both the distractor voices were females of approximately the same age as the matched guise speaker.

The distractor voice for the American-accented recording was from a speaker born in the U.S. and spoke a variety typical of the northwestern part of the country. I included this kind of speaker so that participants would hear a voice that they would likely identify as “American.” The distractor voice for the Philippines-accented recording was from a speaker who had recently migrated to the U.S. and used a variety of English that was common to bilingual speakers of Tagalog and English. I included this speaker with the intention of having a highly proficient English user which would be possibly unfamiliar to participants and offer a noticeable contrast to the other varieties. I included these two voices as a strategy to distract the participants from realizing that the guise speaker had made two of the recordings.

#### Stimuli.

What kind of speech to use as stimuli was carefully considered. In order to minimize the variables to the recordings as much as possible, the chosen stimuli involved having the speakers read from the same text. Two problems with this approach are what Garrett (2010) called the *salience question* and the *style-authenticity question*. The *salience question* points out “that providing respondents with the repeated content of a reading passage presented by a series of voices may exaggerate the language variations and make them much more salient than they would normally be outside the experimental environment” (p. 57). With respect to the *style-authenticity question*, Garrett described how Labov “employed reading aloud as a technique for eliciting a relatively formal style of language in order to contrast its linguistic features with those of more casual or

spontaneous styles.” Garrett cautioned that “it may not be wise to assume that more spontaneous speech will be evaluated in the same way” (p. 59).

As a compromise, the speakers were asked to emulate a TED Talk. The sample was chosen as it was a naturally occurring sample of speech with a context that could be understood by itself, an acceptable level of content and vocabulary, and was originally spoken as part of a public speech, which was appropriate as the participants were asked to evaluate the speakers’ public speaking ability.

I transcribed a 30 second section of a TED Talk and asked the speakers to first become familiar with the transcription and then to record themselves reading it aloud as though they were speaking to an audience. The transcription is provided below:

Take wind energy. It’s one of my favorite examples of perspective shifting. There’s no way that we’re going to build a better standard wind turbine than the experts in that industry. But we found a way to get up higher into the sky, and so get access to faster, more consistent winds, and so more energy without needing hundreds of tons of steel to get there. (Teller, 2016)

In answer to *saliency question* (respondents hearing the same content four times in a row) respondents were asked to pay attention to the speakers as though they were evaluating their public speaking ability. As for the *style-authenticity question*, a TED Talk speech was chosen with the idea that because the original script was spoken, the recordings would simulate a speaking style instead of sounding like someone reading from a script.

A further issue with the chosen stimuli was *the mimicking-authenticity question* (Garrett, 2010, p. 58). This has to do with the matched guise aspect, where it has been found that “there can be many inaccuracies when people are asked to mimic accents, and even if respondents are able to ‘validate’ the voices, they might nevertheless perceive the

voice to be ‘odd’ in some way” (p. 58). Although the JE condition was the more unnatural of the two guises, it was decided that the guise speaker would be able to convincingly mimic a Japanese-accented variety of English due to her familiarity with not only how it sounds but also with what it feels like to actively attempt to change the sound.

A final issue about the audio samples that is worth discussing is *the neutrality question*. “Given the ways in which we interpret texts as we read them, drawing upon pre-existing social schemata, the concept of a ‘factually neutral’ text cannot be assumed to be unproblematic” (p. 59). Although it may be impossible to choose content that is “factually neutral,” the topic – wind energy – was chosen with the idea that it would appeal to a college-educated audience on a general level and with the intention of improving the probability that listeners would focus on evaluating the public speaking ability of the speakers, rather than feel compelled to analyze the message or feel offended, uncomfortable, or anxious about the content.

Further factors affecting listener evaluations.

Previous research on language attitudes has revealed a number of factors that can significantly affect listener evaluations (Garrett, 2010). Factors controlled for in the current study include age and gender – both conditions were of the same speaker so these factors do not vary (this will be discussed later in terms of limits to generalizability); accent strength – the JE condition was of a moderate accent which was intended to sound distinctly foreign but only in terms of its phonology, not lexical or syntactical; rate of speech – all four conditions were monitored for similar speed; setting – as the respondents took the survey online, they were asked to use headphones and/or be in a



quiet environment so that they could clearly hear the samples; and finally left-right bias was avoided by programming the Qualtrics software to randomize which side the positive and negative traits appeared in the semantic differential.

#### The Semantic Differential (SD).

Developed 50 years ago by Osgood, Suci, and Tannenbaum (1969) and still frequently utilized in fields such as information systems research (Verhagen, van den Hooff, & Meents, 2015), the semantic differential is a widely applied measurement technique. The SD “is a technique to measure the psychological meaning of concepts, or a person’s subjective perception of and affective reactions to the properties of concepts through the use of bipolar scales or bipolar items” (Verhagen, van den Hooff, & Meents, 2015, p.110). Each semantic scale “is assumed to represent a straight line function that passes through the origin of this space, and a sample of such scales then represents a multidimensional space” (Osgood, Suci, & Tannenbaum, 1969, p. 63).

One reason behind the decision to use a semantic differential as part of the instrument for this study is the ease with which data can be collected and then analyzed quantitatively. For example, having participants rate speakers on scales that are relevant to the study creates data very differently from having participants listen to different accents and then asking them to answer open-ended questions.

Choosing which bipolar scales to use in the study was carefully considered. As Osgood, Suci, and Tannenbaum (1969) pointed out that “rather than relying on the spontaneous emission of words relating to a particular stimulating sign, we need to play a game of ‘Twenty Questions’ with our subject” (p. 58) and that “the logical tool to uncover these dimensions is factor analysis” (p. 64), the Speech Evaluation Instrument

created by Zahn and Hopper (1985) was adapted for this study. Zahn and Hopper used scales found in other language attitudes studies and used factor analysis to uncover which scales were the most salient to the subjects – undergraduates at two different universities in the United States. As noted above, Zahn and Hopper found 30 bipolar scales and grouped them into three dimensions that they named status (12 item), solidarity (11 items), and dynamism (7 items).

All 30 scales from the Zahn and Hopper study were replicated exactly for the current study, with four exceptions. I changed four of the scales that only added an *un-* morpheme to represent the negative term. I changed “Unintelligent” to “dull,” “uneducated” to “ignorant,” “unfriendly” to “aloof,” and “unlikeable” to “mean.” I made these changes because I felt that there was an overabundance of scales where the only difference between the positive and negative side was the addition of the *un-* morpheme. By changing the word in the scales to a different antonym, I believed that it would engage the participants more in filling the scales and reduce the tendency to choose the center evaluation, indicating a neutral response.

While it is possible that the comparability of this study to similar studies was compromised when I made the choice to change some of the negative terms, I initially adapted Zahn and Hopper’s list for this study with respect to the frequency in which other studies have used the same words to create scales. As such, some discretion may be advisable when making comparisons between this study and other studies that did not change any of the terms.

<b>Status</b>	<b>Solidarity</b>	<b>Dynamism</b>
clear – unclear	honest – dishonest	active – passive
intelligent – dull	sweet – sour	talkative – shy
educated – ignorant	nice – awful	aggressive – unaggressive
fluent – disfluent	good natured – hostile	enthusiastic – hesitant
organized – disorganized	kind – unkind	strong – weak
rich – poor	warm – cold	confident – unsure
complete – incomplete	friendly – aloof	lazy – energetic
literate – illiterate	likeable – mean	
upper class – lower class	pleasant – unpleasant	
White collar – Blue collar	considerate – inconsiderate	
advantaged – disadvantaged	good – bad	
experienced – amateur		

Table 2 – Dimensions and Scales

I also considered whether to use parametric statistics in analyzing the data for this study, assuming normal distribution. This consideration is due to the debate between two views: measurement and statistics:

In brief, proponents of measurement hold that level of measurement (nominal, ordinal, interval, ratio) constrains the kinds of statistical procedures that can be applied to the numerical data. The proponents of statistics maintain that...the level of measurement is not a constraining factor. Those who accept the latter view tolerate the use of parametric statistics with scores from quasi-interval scales that actually are at the ordinal level of measurement, a common practice that is criticized by proponents of the former view. (Dawis, 1987, p. 487)

Language attitude studies frequently ask participants to rank items on Likert scales, which means that the data is ordinal and that it is impossible to be sure that the distance between each possible marking on a scale corresponds to a number in the same way for every participant. For example, “strongly agree” and “somewhat agree” may be assigned a 5 and a 4 respectively, but the difference between “strongly” and “somewhat” might not actually have a different value of exactly one and may even be very different for each

person. As such, proponents of the measurement view would argue that parametric statistics cannot be used for the data.

Taking the statistic view and using parametric statistics reduces the reliability of the study if one holds the view that the numerical values attributed to the scale positions do not reflect reality close enough to make any strong conclusions. On the other hand, the measurement view does not seem to have taken hold in the field of language attitudes as many of these studies (e.g., Bauman, 2013; Tokumoto & Shibata, 2011) have used semantic differential and Likert scales with parametric statistics. Furthermore, Osgood, Suci, and Tannenbaum (1969) asserted that they “have amassed a considerable amount of data on reliability” on their semantic differential instrument and that “for group data (“cultural meanings”), changes or differences in measured meaning as small as one half of a scale unit are significant at the 5 per cent level” (p. 79). As such, while I believe that there are logical arguments against using parametric statistics for data gathered by Likert scales, I decided that there is enough support for the other side, notably that it seems to be common practice in language attitude research. Thus, I used parametric statistics once the data was shown to be normally distributed. I discuss the weaknesses of this approach further in the Discussion chapter.

The validity of the SD can be determined by whether the instrument measures what it is supposed to measure. Developers of the SD, Osgood, Suci, and Tannenbaum (as cited in Brown, 1992), believe that the instrument “organizes ideas and distinguishes between concepts in a manner that approximates ordinary decision-making processes” (p. 48).

The SD has face validity when the researcher has made good judgement in selecting stimulus items for participants to respond to and when the items are relevant to the area of research interest. This is where the SEI developed by Zahn and Hopper (1985) was useful. As a caveat, it should be noted that the SEI has been criticized (Garrett, 2010), and some of its shortcomings are discussed in the Discussion chapter. Even so, through Zahn and Hopper's research, which resulted in easier comparisons across studies, the SEI has been adapted and adopted by many language attitude studies. As the current study will be measuring language attitudes using very similar methods to other such studies, it can be said that the SEI and the SD possess face validity.

In terms of reliability, the SD can be said to fulfill this criterion if participants arrive at the same scores were they to complete the survey more than once. Upon readministration of the SD, the developers "found that there was such close agreement between items that the variances approach zero and computed reliability coefficients become meaningless" (Osgood, Suci, & Tannenbaum, 1975, as cited in Brown, 1992).

Owing to the research carried out by Zahn and Hopper (1985) to develop the SEI, as well as the testing of the SD by Osgood, Suci, and Tannenbaum (1969), it can be argued that the chosen instrument for this study adheres to general reliability and validity parameters.

## Results

This results section is divided into two parts. In the first part I discuss the mean scores of both accent conditions in the three dimensions, the statistical operations that were used to determine normal distribution of the data, the results of testing for interaction between the groups that listened to the speakers in different order, and the results of tests for statistical significance between ratings for the two guises on each of the three dimensions.

In the second part of this section, I discuss the research question and hypothesis in view of the results of the data, followed by a discussion of the results from the survey question “Where do you think this speaker is from?” which respondents were required to answer after each recording. I finish this section with a brief description of which of the characteristics had the biggest differences between ratings of the guises.

The first step of the analysis was to determine if the data were normally distributed so that it could be determined whether parametric tests were appropriate for running the data. According to Kolmogorov-Smirnov and Shapiro-Wilk tests, the data were mostly normally distributed ( $p > .05$  only for the Status dimension for the group that heard JE first and AE second). As such, it was decided to run parametric tests on the data.

As described above, the raw scores were adjusted so that a rating closer to the negative term in a bipolar scale always reflected a more negative evaluation and a rating closer to the positive term always reflected a more positive evaluation (on a scale of 1 to 5). Mean scores (standard deviations in parentheses) by accent condition and dimension are given in Table 1.

	Status	Solidarity	Dynamism
<b>American-accented English</b>	3.72 (.635)	3.84 (.689)	3.04 (.603)
<b>Japanese-accented English</b>	3.45 (.661)	3.66 (.672)	2.83 (.741)

Table 3 – Mean scores and standard deviations by accent condition and dimension

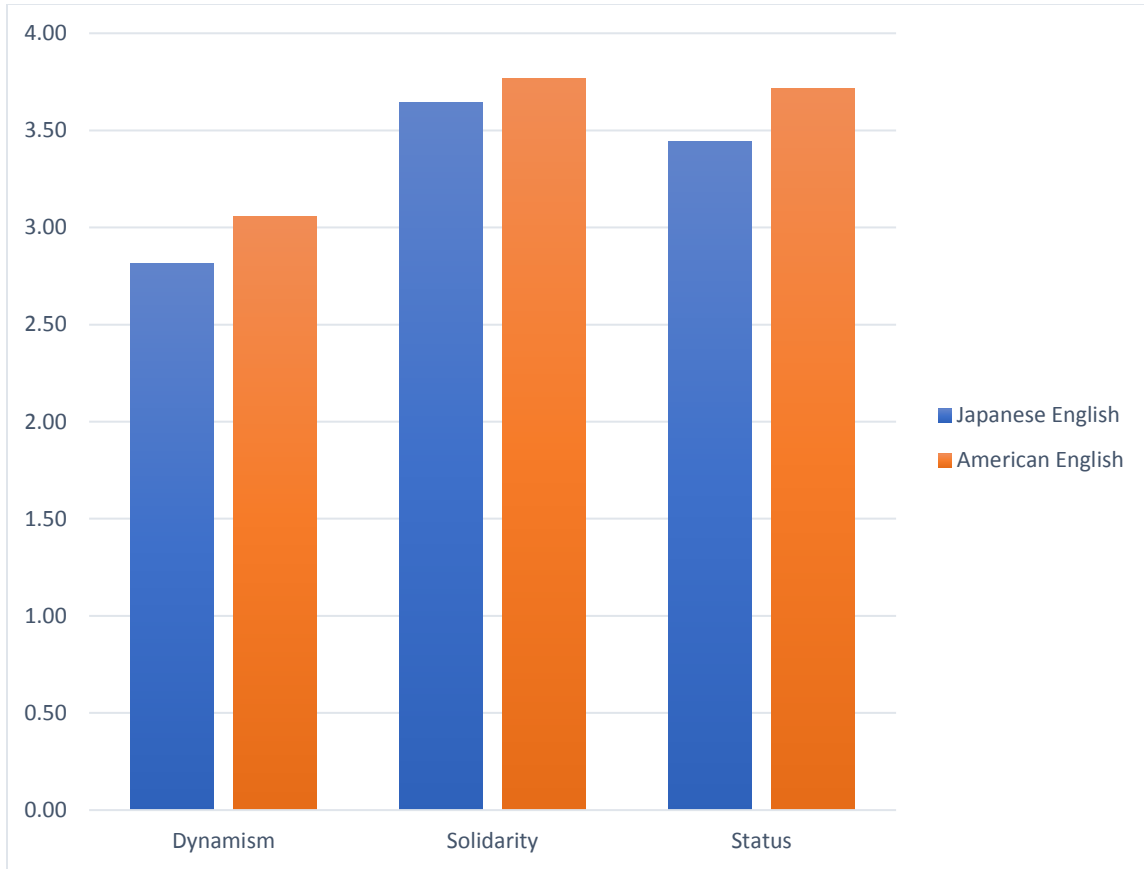


Figure 1 – Ratings by Dimension and Language Condition

The AE condition was rated higher than the JE condition on all three dimensions. To test whether these differences were statistically significant, I performed repeated measures ANOVAs on each of the dimensions for the within-subjects variable (i.e., the AE/JE conditions), as well as the between-subjects variable (i.e., the order in which the conditions were heard). In all statistical results reported below, a significant result was defined as a *p* value of less than .05.

The effect of the accent condition on the Status dimension was significant,  $p = .022$  ( $<.05$ ) with a partial eta squared of .141, indicating a small effect size. Ordering was not found to be significant with  $p = .553$  ( $>.05$ ). Interaction between accent condition and ordering was also not found to be statistically significant with  $p = .183$  ( $>.05$ ).

The effect of the accent condition on the Solidarity dimension was significant,  $p = .04$  ( $<.05$ ) with a partial eta squared = .115, indicating a small effect size. Ordering was not statistically significant with  $p = .382$  ( $>.05$ ). Interaction between accent condition and ordering was also not found to be statistically significant with  $p = .649$  ( $>.05$ ).

For the Dynamism dimension, the effect of the accent condition was not found to be statistically significant ( $p = .069$ ). Ordering was  $p = .865$ , indicating no statistically significant difference. The interaction between accent condition and ordering was also not found to be statistically significant with  $p = .676$ .

In summary, across all three dimensions, there was not a statistically significant difference between the ratings by Group 1 ( $n = 18$ ), where participants heard the American guise first and the Japanese guise last, and Group 2 ( $n = 22$ ), where the Japanese guise was first and the American guise last. Additionally, interaction between order and accent condition was not found to be statistically significant. These results are evidence that in this experiment, at least, the order in which participants heard the recordings did not affect the evaluations at a statistically significant level. Finally, although no statistical significance was found between the guises for the Dynamism dimension, the AE condition was rated significantly more favorably than the JE condition on the Status and Solidarity dimensions, although the effect sizes for both were small.



In light of the data it is possible to answer the research question: Will the guise of a female Japanese-accented English speaker in her twenties be evaluated negatively on traits related to solidarity and dynamism, but more favorably or equally on status-related traits when compared with evaluations of the same speaker in an American-accented English condition?

The simple answer is that the hypothesis was shown to be untrue: In comparison to the AE guise, the JE guise was rated unfavorably in both the Status and Solidarity dimensions. However, even though the JE guise was rated lower on these dimensions consistently enough to be significant, because of the small effect size, it is not unreasonable to surmise that the actual attitudes held by participants about JE might not be noticeable in a normal, everyday context.

“Where do you think this speaker is from?”

As discussed in the methodology, a Matched Guise was chosen over a Verbal Guise due to the availability of a speaker who was ostensibly capable of using both Japanese-accented English and American-accented English. Testing of the ability of the Matched Guise speaker to use both varieties naturally was conducted by asking participants to answer the question “Where do you think this speaker is from?” after listening to each of the recordings. A summary of the results from this survey question are represented below in Table 4 and Table 5.

<b>American English Condition</b>	<b>Number of guesses</b>
United States (if including Canada)	25 (27)
Latin America or Mexico	4
Asia	3
“no idea”; “not sure”	2
Europe	1
Japan	1
Malaysia	1
“Non-native speaker. Non-count noun issues. American West Coast”	1

Table 4 – Speaker Origin Survey Question for American English

In the American English condition, 27 out of 40 participants guessed that the guise speaker was from an English-speaking country, 7 participants provided answers which suggest that they were able to detect that the speaker was not American (but also not from Japan or an Asian country), 3 were able to hear something in the speech that sounded “Asian” to them, 2 gave noncommittal answers, and 1 assessed the speaker as Japanese.

<b>Japanese English Condition</b>	<b>Number of Guesses</b>
Japan	24
Asia, East Asia, SE Asia	11
Japan, Korea, or China	2
Japan or South Korea	1
India	1
Portland	1

Table 5 – Speaker Origin Survey Question for Japanese English

In the Japanese English condition, 24 out of 40 participants were able to correctly identify the variety as one spoken in Japan, 14 participants thought that the speaker sounded either “Asian” or listed multiple countries associated with Asia, 1 wrote India, and 1 listed Portland.

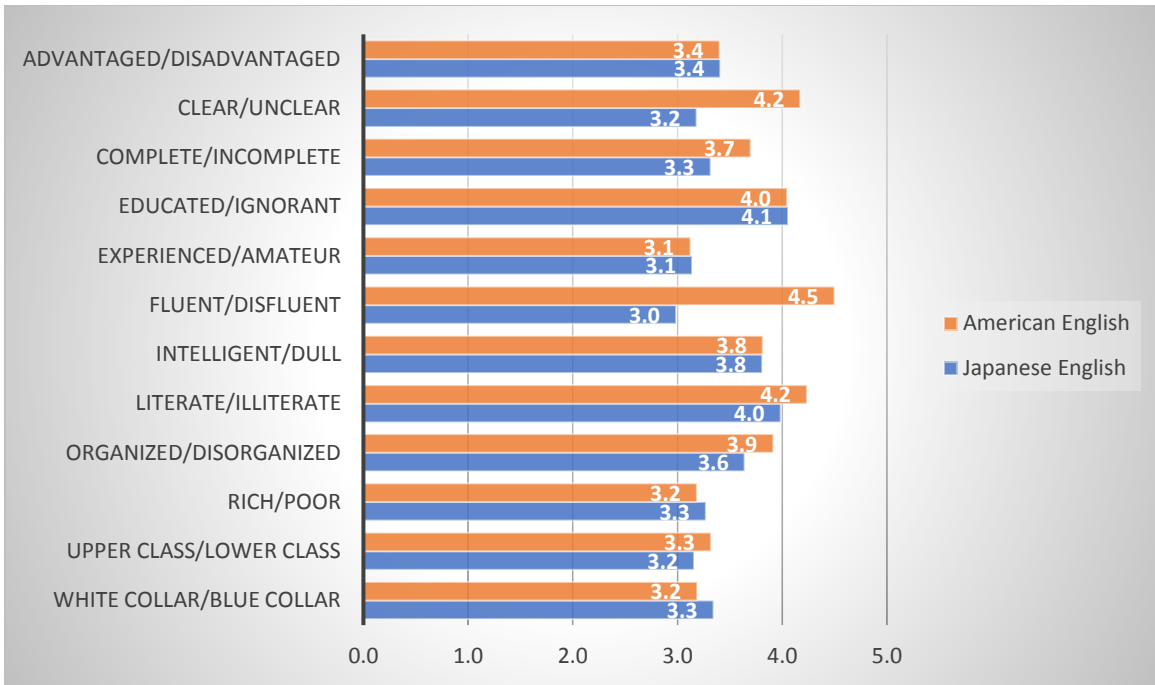


Figure 2 – Status Dimension Ratings by Individual Scale

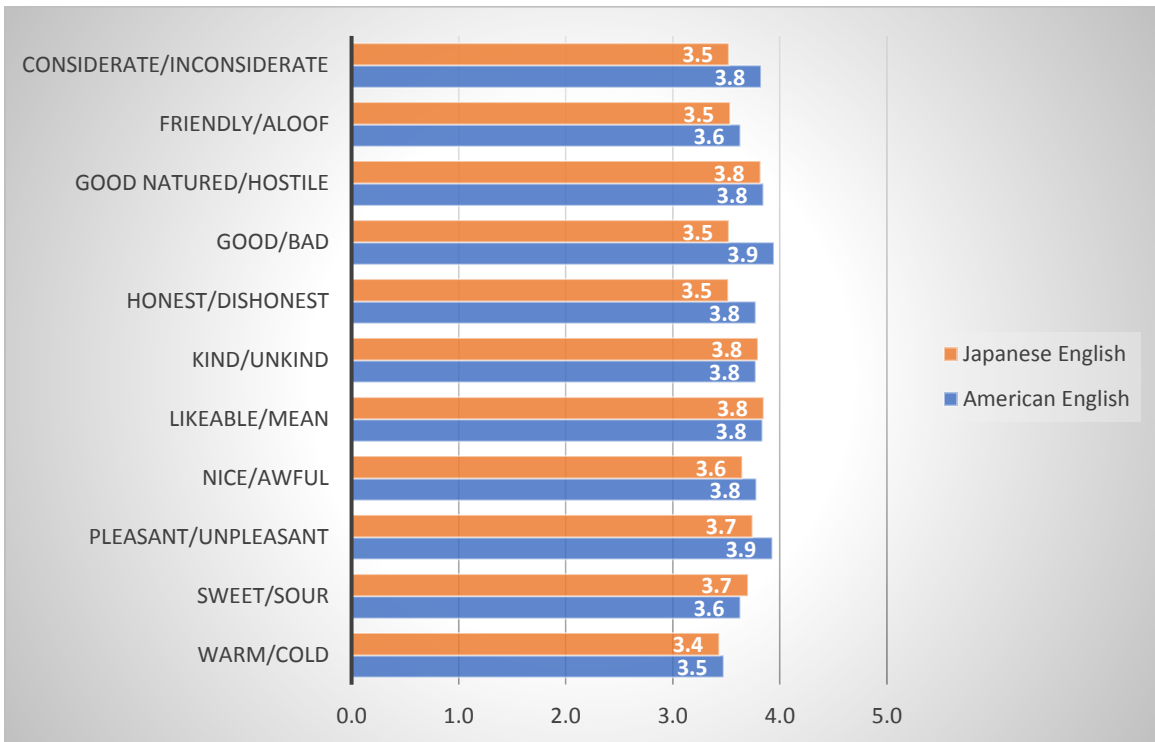


Figure 3 – Solidarity Dimension Ratings by Individual Scale

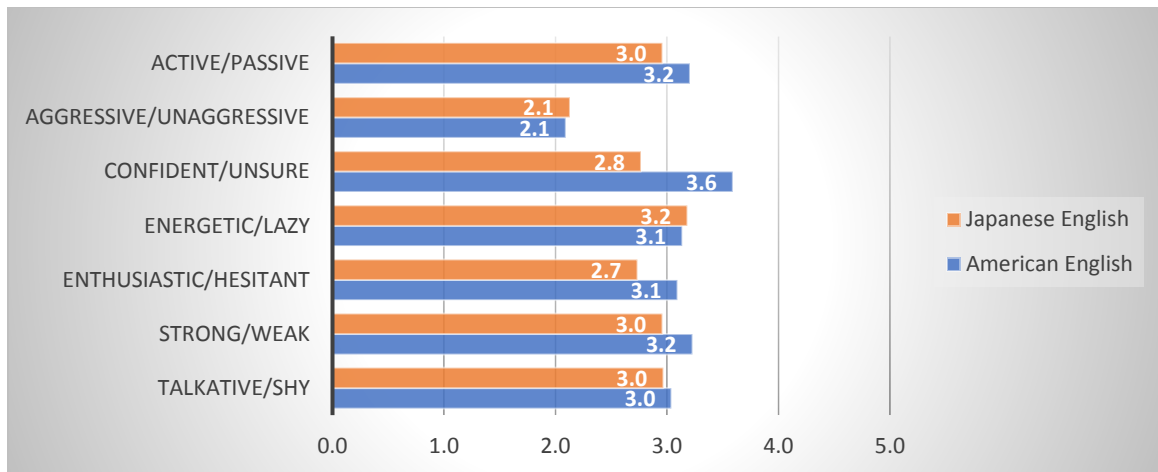


Figure 4 – Dynamism Dimension Ratings by Individual Scale

Figure 2, 3, and 4 are organized with all traits in each dimension of Status, Solidarity, and Dynamism, respectively. Traits are listed on the left with the positive end of the scale listed first and the negative second (e.g., Advantaged/Disadvantaged). Higher scores represent a more positive evaluation. As Figure 2, 3, and 4 illustrate, several characteristics differed by only 0.2 or 0.1, and some did not appear to differ at all. Interestingly, the JE guise was rated by 0.1 higher on five of the scales (White collar/Blue collar, Sweet/Sour, Rich/Poor, Energetic/Lazy, Educated/Ignorant). On the other hand, there were a few characteristics where the AE guise was rated noticeably higher. The AE guise was higher by 0.4 or more on Good/Bad, Enthusiastic/Hesitant, Confident/Unsure, and Complete/Incomplete. The AE guise was higher by a whole point on the Clear/Unclear scale and by 1.5 on the Fluent/Disfluent scale.

Also revealing from Figure 4 is the average rating for the Aggressive/Unaggressive characteristic. Although there was apparently no difference between language conditions, the low rating of 2.1 is lower than on any other scale by 0.6.

## Discussion

As stated in the title of this study, it was my intention to conduct a language attitudes study within the World Englishes paradigm. While carrying out such a study in this way might only minimally affect the methodology, data, and results, interpretations of the data and results are potentially affected in major ways. I begin this Discussion section by making comparisons between the study that I conducted and the Cargile and Giles (1998) study, which was very similar to mine both in the language varieties that were examined as well as methodologically.

An important difference is that whereas the Cargile and Giles study examined Japanese-accented English as a variety of language used by a culture which might pose an economic threat to the “native speaker” participants living in the United States, I envisioned Japanese-accented English as a variety of English spoken in one of Kachru’s Expanding Circle countries which was to be evaluated by proficient users of English living in Portland, Oregon.

From a World Englishes perspective, I am working from Kachru’s (1986) idea that “English belongs to all who use it.” Further, scholars that subscribe to the World Englishes paradigm have conducted research showing that “one’s English is more comprehensible to those people who have had active exposure to it” (Smith & Bisazza, 1982, p. 269), nonnative speakers of English outnumber native speakers (Crystal, 2003; Kachru, 1986), and that “native speakers” are not necessarily the most comprehensible English users (Smith & Rafiqzad, 1979, p. 380). To put it another way, English as a global language only starts to make sense when looking at it globally. Rather than splitting English users in “native speakers” and “nonnative speakers,” placing them into

Kachru's Inner, Outer, and Expanding Circles, where each circle is permeable (i.e., an Expanding Circle region can potentially move to the Outer Circle under the right circumstance), provides at least a different starting point for what might be assumed about English users from certain countries.

It is from this point that I try to interpret the data from my study. After comparing results between my study and the Cargile and Giles study I discuss other aspects of World Englishes as well as other perspectives which might be worth considering.

The hypothesis of this study predicted that a Japanese-accented variety of English would have a linguistic profile similar to a high prestige variety such as American or British English, as found in the Cargile and Giles (1998) study. Specifically, the hypothesis stated that listeners would rate Japanese-accented English similar to American-accented English on all three dimensions: Status, Solidarity, and Dynamism. Analysis of the data showed that participants rated Japanese-accented English lower than American-accented English on the Status and Solidarity dimensions but not on the Dynamism dimension at a statistically significant level. In terms of the *p* value in statistical analysis, then, the hypothesis was not supported. These results would seem to contradict the findings of the Cargile and Giles (1998) study, which contended that "Japanese-accented English is a special variety of non-standard speech in America" (p. 342) where "American images of the Japanese suggest that they also have high standing, but may encourage few warm feelings" (p. 342).

On the other hand, even though the ratings were shown to be significantly lower on two out of three dimensions, due to the small effect sizes, these results may not necessarily contradict the unique evaluative profile of Japanese-accented English. Cargile

and Giles' suggestion of a unique profile for Japanese-accented English was based on comparisons with other previous research that used American listeners, which had shown that Outer or Expanding Circle English users either fall into a group that is evaluated less favorably on Status and equally or more favorably on Solidarity or into a group that is evaluated less favorably on both Status and Solidarity when compared to Inner Circle speakers (pp. 350-351). For Cargile and Giles, Japanese-accented English did not fit into either of these groups because their data demonstrated that "a (non-standard) moderate Japanese-accented speaker compares favorably on status-related traits, but unfavorably on traits related to attractiveness" (p. 349, parenthesis in original). In view of the data for the current study, the *p* values indicated that the participants in this study consistently evaluated Japanese-accented English less favorably in the Status and Solidarity dimensions, but the small effect sizes suggest that the differences might be negligible. In other words, the data suggests that the linguistic profile of Japanese-accented English might fit into the group of Englishes that is evaluated less favorably than American-accented English on both the Status and Solidarity dimensions but that the amount that it is rated down is rather small.

"Where do you think this speaker is from?"

After listening to each of the four recordings, the respondents were asked the question, "Where do you think this speaker is from?" before continuing with the next part of the survey. The question was included in response to Garrett's (2010) suggestion that respondents be asked where they think a speaker is from as a means of validating the guises (p. 58). Other studies (Cargile & Giles, 1998; Lindemann, 2003) included a parallel question as well for similar reasons.

The ideal result for this question in response to the AE condition would be for all of the participants to state that the speaker was from the U.S., or possibly Canada. The number of respondents that wrote in a variation of either “America” or “Canada” was 27 out of 40, or approximately two-thirds. There are two ways to view this result. One is that the speaker, who started trying to sound like an American when she moved to the U.S. at the age of 12, impressively made 2 out of 3 proficient English users unaware that they were listening to a bilingual whose first language was not English. The other way to interpret this result is that 1 out of 3 people detected something in the speech that led them to believe the speaker was not from an Inner Circle country. From this perspective, it might be useful to consider the matched guise technique as more difficult to conduct, or more flawed, than some language attitude researchers have wanted to believe.

Cargile and Giles (1998) also faced challenges in validating their matched guise speaker as a user of American-accented English. For example, when they were pilot testing their American-accented English with a “non-aggressive” passage, “77% of listeners reported the speaker’s background as American” (p. 345). As a rough comparison to the present study, the Cargile and Giles study achieved only about 10% better results. Apparently even more challenging, however, was to have the same speaker read an aggressive passage, as “initially only 10% of listeners reported the speaker’s background as American” (p. 345). After several re-recordings, “a version was selected which led to a slight majority (53%) of pretest participants to report the speaker’s background as American” (pp. 345-346).

Data from both this study and the Cargile and Giles (1998) study does not seem to offer strong support in favor of the matched guise technique, calling into question its



construct validity. Even so, the fact that the AE condition received higher evaluations than the JE condition at a statistically significant level is undeniable evidence that the participants responded more favorably to speech that either more closely approximated American-accented English or at the very least did not sound foreign (i.e., “not American”).

The results for the question being asked about the JE condition are also important but for different reasons. The question operates here as means to test how many respondents could correctly identify Japanese-accented English as such, as well as how many would list other Asian countries, use the broader term “Asian,” or name some other geographical location. The result of 24 out of 40 (60%) correctly guessing that the speaker was from Japan was a surprisingly high amount when compared with other studies. In the Lindemann (2003) verbal guise study, for example, where the “non-native speakers” were Korean, only 8% of the respondents – undergraduates at the University of Michigan – successfully guessed “Korea” (p. 354). The results from the speaker origin question in the present study were not without problems, however, which are discussed in the limitations section.

Setting aside the discussion about how the participants in the present study were able to identify Japanese-accented English, what it means when participants can or cannot identify an English variety is worth consideration. A respondent correctly guessing “Japan,” for example, allows the opportunity to put them into a group that made evaluations based on the actual variety in question: Japanese-accented English. Additionally, respondents guessing “Asian” or listing other Asian countries might represent a group that lacks the ability to differentiate between these varieties but

nonetheless associates what they heard with a certain area of the world. Yet another group might consist of respondents who confuse the speech with a group of people from an entirely different geographical location. By conducting analyses on these different groups, it would be possible to determine if a groups' ability to identify an English variety or not is an important variable that affects how respondents make their evaluations.

#### Ratings by individual characteristic

While a comprehensive analysis of how participants responded to each individual characteristic was not part of the goal of this study, the results in Figures 2, 3, and 4 merit at least a brief discussion here and are further considered below regarding future research.

Figures 2, 3, and 4 make a few things easily apparent from the data. One is that many of the characteristics received mean ratings between 3.0 and 3.8 and that a considerable number differed between the guises by either 0.2 or less. The scales for Talkative/Shy, Strong/Weak, Rich/Poor, Experienced/Amateur, Energetic/Lazy, and Active/Passive, for instance, all received mean scores between 3.0 and 3.3 for both guises. This would seem to suggest that respondents had a neutral attitude on these characteristics toward the speakers. Conversely, the two characteristics with the biggest difference in mean ratings were Fluent/Disfluent (AE was higher by 1.5) and Clear/Unclear (AE was higher by 1.0). Since it is not possible to determine from the data how important these characteristics were to the respondents, it is difficult to say what the consequences might be for these differences in ratings. What can be stated, however, is that respondents felt markedly different about the two guises in regards to these two

characteristics and that if these attitudes carried over to behavior, then they could potentially have a negative impact on a user of Japanese-accented English.

Also revealing from Figure 4 is the mean ratings for the Dynamism characteristic Aggressive/Unaggressive, where “Aggressive” was designed to be on the positive side of the scale. There was no difference between language conditions, but the 2.1 ratings were lower than on any other scale by 0.6 or more. Garrett (2010) suggested that “a high rating on a semantic differential scale for a particular trait does not necessarily indicate favorableness” (p. 71). For the Aggressive/Unaggressive scale, the situation appears to be the reverse, where Aggressive was not perceived to be a positive trait for the speaker in either of her guises. This became evident in some of the surveys when respondents would give positive ratings on every item in the Dynamism scale except for Aggressive/Unaggressive. Garrett argued in favor of having respondents rank traits according to the supposed context:

This practice of asking respondents how much value they place on the various traits would seem to have high utility, especially if such values could differ across groups or contexts where, for example, ‘honesty’ might be seen as of more significance than, say, ‘liveliness’ in one context, but of less significance in another. (p. 71)

For the context of the current study, it seems reasonable to suggest that aggressiveness is not a positive trait for a female speaking publicly about wind energy. This last statement might simply be stating the obvious, but it also provides evidence that the selection of scales for this study might have benefited from more individual scrutiny.

### Limitations

The response rate from the online survey was relatively low ( $n = 40$ ), which limits the generalizability of the findings. Even so, a statistically significant difference in ratings

was found between the guises for the Status and Solidarity dimensions. That a difference was found with a small sample might be indicative that a larger sample and further analysis would reveal more interesting results.

Generalizations that can be made about the varieties spoken by the guise speaker are likely limited to female Japanese-accented English and American-accented English users in their twenties, as both age and gender of the speaker has been shown to affect listener evaluations (Garrett, 2010). It might be unreasonable to expect similar results if a male in his fifties were to perform the guises, for example.

Furthermore, with respect to the methodological choices of using a script for the recordings, as well as having one speaker talk in two separate guises, it might be reasonable to suspect that the participants were not responding to one voice (i.e., the guise speaker) talking about wind turbines. In other words, because each of the recordings was the same, scripted content, the listeners were led to assume that the speakers did not create the content themselves. There was some disparity between the guise recordings as well. Such as the Japanese-accented English recording being about 10 seconds longer than the American-accented English recording. With different speech rates, it is possible that participants did not respond to the two guises as intended. That is, two similar people speaking on the same topic in different varieties of English. These two factors possibly affected evaluations in unexpected ways.

As for the listeners, the population was a convenience sample which represented college-educated proficient users of English living in the Portland area. While this population likely represents the kind of people who would likely have interactions with

either of the personas represented by the guise speaker, caution would be appropriate in making any generalizations toward different groups of people.

Also pertinent to the convenience sample was the relatively high amount of participants who were able to correctly identify the Japanese-accented English guise as a Japanese person – 24 out of 40. This result suggests that the Japanese-accented English provided in this study was a familiar prototype to a majority of the sample population. However, as minimal personal information was gathered from participants, such as exposure to Japanese speakers of English, some caution is necessary when interpreting these results. One possibility is that a large portion of the participants were people whom I, the researcher, knew personally. Since these participants would know about my interest in the Japanese language, culture, and people, and some of them were likely interested in Japanese culture themselves, it is not improbable that some of these participants used this knowledge while identifying the speakers.

Additionally, due to the level of anonymity allowed in taking the survey, where participants were not required to supply any form of contact information, it was impossible to control many aspects of the data, such as making sure that nobody took the survey more than once. Follow-up interviews were also impossible to conduct. For example, I would have liked to talk to the participant who guessed that the speaker for the AE condition was from Japan. Were they able to hear something in the speech that led them to conclude that the speaker was Japanese? Could they tell that the JE and AE guises were the same person? Or might they have recognized the speaker and therefore known that the speaker was Japanese just by hearing the voice? I could not ask these types of questions of respondents because I had no way of knowing whose results I was

looking at. The anonymity allowed to participants did make the survey easy to disseminate, however.

Some of the problems mentioned above could be resolved by pilot testing and/or making small improvements to the survey, such as adding optional questions and comment sections. The issue with participants potentially repeating the survey could be controlled by a function in Qualtrics which restricts the number of times the survey can be taken from one location. This function was not employed for this study, however, because of the potential for numerous people in one location (i.e., IP address) to be participants. Enabling the function may be a more prudent choice for future studies.

#### Implications for future research

Utilizing the matched guise technique resulted in the exclusion of other varieties of English, limiting the comparison between only two. For example, it would have been interesting to learn how the guise speaker's Japanese-accented English and American-accented English compared to the Philippines-accented English and the other American-accented English distractor voices. I did not analyze the data I gathered from the distractors because I followed the example of previous research (e.g., Lindemann, 2003) where distractor voices were used but their data not included in the analyses. This is but one example of the kind of potential herd mentality in language attitude research (or any research), where researchers replicate methodologies from other studies when they observe that everyone else is doing it, so they conclude that the methods must be valid. While it may be true that using data from the distractor voices would stop a Matched-Guise study from representing its original form, this can always be noted and the reader can decide how to interpret the findings.

Furthermore, the matched guise speaker was apparently able to convince only two out of three participants that she spoke American-accented English. While this is an impressive feat, two out of three is not a scientifically sound ratio. On the other hand, the matched guise technique made it possible to control many potentially confounding variables, such as voice quality, pitch, volume, and rate of speech, which are more difficult to manipulate in a verbal guise study. As a result, the data provided some picture of how much attitudes can change and/or stay the same when the only manipulated variable is the language variety. Future studies may have better data if researchers can manipulate the recordings to better represent the language varieties in question. One such solution may be to use a verbal guise, though this would create a new set of problems.

Another challenge with the matched-guise technique is setting up the survey in a way that prevents participants from focusing on foreign-sounding language. The current study was flawed in this regard. A notable design might be attributed to the survey question, “Where do you think this speaker is from?” The question brings the participants’ attention to the fact that they might be listening to speakers from a different place. Future studies may benefit by following Lindemann’s (2003) methodology, which only asked participants to assess the ethnicity of the speakers after they had completed all of their other evaluations.

As for the SEI, when Zahn and Hopper (1985) created theirs 30 years ago, they stated that it was “one instrument which shows promise as a valuable measure of language attitudes” (p. 121). While it may be true that their SEI was a step forward for language attitude research, there are notable risks involved in adopting the characteristics

that Zahn and Hopper found to be the most salient in their study. As Garrett (2010) explained:

The advantage of using labels from previous studies is that it can save a great deal of time, and allow a reasonable degree of confidence that one has covered the main evaluative dimensions along which respondents are likely to be making their judgements. It may also allow better comparability across studies. The difficulty with it is that it may lead to some circularity in which the same dimensions simply become better documented and so are assumed to be exhaustive. (p. 56)

Zahn and Hopper themselves suggested that their SEI was only a preliminary step towards systematizing language attitudes research. As they wrote in their study:

We recommend that extension of speech evaluation research to new speech communities and contexts include interview and ethnographic assessment of evaluators concerns that may not be directly reflected in the items of the SEI. Bipolar adjectives obtained from such research concerning the particular language situation or context might be profitably combined with items from a shortened version of the SEI. We also recommend the items in the SEI be subjected to factor analysis in each study of speech evaluation. (p. 121)

Research on language attitudes, including this one, often seem to skip the step of figuring out ways to elicit characteristics that are salient to the linguistic context within the studies. As such, keeping all 30 scales was probably unnecessary. Conducting pretests and pilot studies would have been helpful in deciding which scales to keep. Reducing the number of scales by half, with 5 of the most salient words for each dimension might have been a more reasonable number of characteristics to evaluate. On the positive side, being able to see how participants rated the guises on 30 different scales provided a considerable amount of data and revealed that many of the characteristics received very similar ratings on both guises and very different ratings on a few, as discussed above.

Another useful addition that future researchers might want to consider would be to adopt or adapt components of the methodology that Lambert et al. (1960) used in their



early study. As discussed in the methodology above about language attitude studies, there is a direct method, which examines the cognitive (or conscious) part of attitudes, and an indirect method, which examines the affective (or unconscious) component. Lambert's study used both direct and indirect methods, and in looking for correlations, found a discrepancy between participants' cognitive and affective attitudes towards the languages in question (English and French in Canada). If utilized in future studies, this kind of methodology could reveal interesting findings regarding conscious and unconscious attitudes towards language.

Also to consider in future research would be how evaluations may have been affected by social desirability bias, where people "give answers to questions in ways that they believe to be 'socially appropriate,'" (Garrett, 2010, p. 44). Due to the "factually neutral" topic for the recording and the manipulated context – evaluating speakers on their public speaking ability – the respondents did not have a salient reason to be invested in different varieties of English because there were no serious consequences for not being able to understand the speakers. Research on international graduate assistants being placed in classes with undergraduates has revealed some strong biases held by the undergraduates who are put in such a situation (Brown, 1992; Rubin, 1992). Applied to a language attitude study, where international graduate assistants are the "guises" and undergraduates are respondents who participate (hypothetically) as students who might be taking classes with the graduate assistants, could set up a context where respondents would be more critical of different varieties of English.

As discussed in the methodology chapter, there are some researchers who argue that ordinal data, such as data gathered by semantic differential Likert-type scales, should

be analyzed with nonparametric statistics since the values between the ratings are unknown. Due to the prevalence of similar language attitude researchers treating their data as parametric statistics, however, I decided to follow their example for my study. Future language attitude studies may benefit in terms of validity either by using nonparametric statistics or by providing arguments for why the data is being treated as a interval or ratio scales.

#### Implications for TESOL

The findings from this study add data to the field of language attitudes, specifically the linguistic profile of Japanese-accented English. The implications of this data might differ depending on whether or not the results are viewed from a World Englishes perspective.

From an unmarked perspective (i.e., non-World Englishes), one may interpret the data as evidence that speaking English with a noticeable Japanese accent results in significantly less favorable judgements than speaking English in a way that more closely approximates a variety of English spoken in America on the Status and Solidarity dimensions. The small effect size may indicate that the degree of difference was not that great, but when it came to judgements on the scales of Fluent/Disfluent and Clear/Unclear, respondents gave markedly lower ratings for the Japanese-accented English guise than for the American-accented English guise.

An implication of this interpretation for Japanese learners of English is that they might benefit from focusing on their pronunciation and trying to emulate a target variety used by “native speakers.” Although they do suggest caution in generalizing from the

results of their data, Cargile and Giles (1998) seem to take their implications a step further:

Japanese face a deficit in terms of being liked by Americans when they speak in English with either a strong, or even a moderate, accent. Thus, for example, if a Japanese figure (e.g., a musician, or a sports hero) wants to encourage popularity among Americans, he or she may be served best by fewer rather than more opportunities to speak English in public. (p. 352)

Taken in this light, data gathered from language attitude studies focusing on Japanese-accented English would seem to indicate that having a Japanese accent is something that should be fixed as soon as possible and Japanese English learners should make this their priority.

At this point it may be beneficial to reinterpret the same results from a World Englishes perspective. My suggestion on how to interpret the data is that Japanese people would benefit from knowing how speakers react to Japanese-accented English, giving them agency to adjust their speech according to how they wish to be perceived. It is also important to consider the possibility that a sample of English users defined by a different set of parameters (i.e., not just “college educated proficient users) or speakers in a different region might react to Japanese-accented English very different than those who participated in this study. Also crucial, however, is that judgements made by Inner Circle speakers of an Expanding Circle variety are absolutely not the most important opinions to consider. As Smith (1983), a World Englishes scholar, stated:

A Thai doesn't need to sound like an American in order to use English well with a Filipino at an ASEAN meeting. A Japanese doesn't need an appreciation of a British lifestyle in order to use English in his business dealings with a Malaysian. The Chinese do not need a background in western literature in order to use English effectively as a language for publications of worldwide distribution. The political leaders of France and Germany use English in private political discussions but this doesn't mean that they take on the political attitudes of

Americans. It is clear that in these situations there is no attempt for the user to be like a native speaker of English. (p. 7)

Looking at the results from this perspective might seem to invalidate the data gathered for this study, but this is not necessarily the case. While there were notable problems with the methodology used in this study, including issues about the matched guise technique, the questionnaire, and the use of parametric statistics where non-parametric would have been more appropriate, the fact that the sample population did not react drastically different to the two guises is still informative. The guise speaker spoke in two markedly different varieties of English, but just as important is the fact that both of the varieties were intelligible. What this means is that even if a speaker uses a very marked accent, if the speech is understandable to the listener, there is the possibility that the negative reactions to the speech will be less than expected.

Gluszek and Dovidio (2010) discussed the challenges of learning a second language and speaking with a “nonnative accent:”

People who are fluent in a second language often speak with a nonnative accent, even after many years in a host country, because they retain the phonology (including intonation) of their native language even when they achieve near-perfect control over other features of the nonnative language. (p. 215)

This might be a fitting description of the guise speaker, who was highly proficient in English and achieved “near perfect control over other features,” but a majority of the participants in the present study still assessed her ethnicity as not someone from an Inner Circle country. The way that I would prefer to interpret these results, in general as well as for English teaching, is that the data offer evidence to support a language learning approach which focuses more on intelligibility and reduces emphasis on sounding exactly like an Inner Circle speaker of English.

## References

- Baugh, J. (2003). Linguistic profiling. In S. Makoni, G. Smitherman, A.F. Ball, & A.K. Spears (Eds.), *Black linguistics: Language, society, and politics in Africa and the Americas* (pp. 155-168). London: Routledge.
- Bauman, C. (2013). Social evaluation of Asian accented English in the United States. *Newcastle Working Papers in Linguistics*, 19(1), 22-41.
- Brown, K. (1992). American college student attitudes toward non-native instructors. *Multilingua – Journal of Cross-Cultural and Interlanguage Communication*, 11(3), 249-266.
- Cargile, A.C. & Giles, H. (1998). Language attitudes toward varieties of English: An American-Japanese context. *Journal of Applied Communication Research*, 26(3), 338-356, DOI: 10.1080/00909889809365511
- Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes toward English accents. *World Englishes*, 14(1), 77-86.
- Chien, S. (2014). Varieties of English: Taiwanese attitudes and perceptions. *Newcastle and Northumbria Working Papers in Linguistics*, 20, 1-15.
- Crystal, D. (2003). *English as a global language* (2nd ed.). New York: Cambridge University Press.
- Edwards, J. (2009). *Language and identity: An introduction*. United Kingdom: Cambridge University press.
- Fasold, R. (1987). *The sociolinguistics of society*. Malden, Massachusetts: Blackwell Publishers Inc.
- Garrett, P. (2010). *Attitudes to language*. United Kingdom: Cambridge University Press.

- Gill, M.M. (1994). Accent and stereotypes: Their effect on perceptions of teachers and lecture comprehension. *Journal of Applied Communication Research*, 22, 348-361.
- Gluszek, A. & Dovidio, J.F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14(2), 214-237,  
DOI: 10.1177/1088868309359288
- Hosoda, M., Stone-Romero, E.F., & Walter, J.N. (2007). Listeners' cognitive and affective reactions to English speakers with standard American English and Asian accents. *Perceptual and Motor Skills*, 104, 307-326.
- Kachru, B.B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H.G. Widdowson (Eds.), *English in the world* (pp. 11-30). Great Britain: Cambridge University Press.
- Kachru, B.B. (1986). *The alchemy of English*. Chicago: University of Illinois Press.
- Kachru, B.B. & Nelson, C.L. (2001). World Englishes. In A. Burns & C. Coffin (Eds.), *Analysing English in a global context: A reader* (pp. 9-25). New York: Routledge.
- Lambert, W.E., Hodgson, R.C., Gardner, R.C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60(1), 44-51.
- Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, 7(3), 348-364.
- Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination*

- in the United States* (2nd ed.). New York: Routledge.
- Matsuda, A. (2003). The ownership of English in Japanese secondary schools. *World Englishes*, 22(4), 483-496.
- Matsuda, M. (1991). Voices of America: Accent, antidiscrimination law, and a jurisprudence for the Last Construction. *Yale Law Journal*, 100(5), 1329.
- Matthews, P. (1997). *The concise Oxford dictionary of linguistics*. New York: Oxford University Press.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1969). The measurement of meaning. In J.G. Snider & C.E. Osgood (Eds.), *Semantic differential technique* (pp. 56-82). Chicago: Aldine Publishing Company.
- Paikeday, T.M. (1985). *The native speaker is dead!* Ontario, Canada: Paikeday Publishing Inc.
- Podberesky, R., Deluty, R.H., & Feldstein, S. (1990). Evaluations of Spanish- and Oriental-accented English speakers. *Social Behavior and Personality*, (18)1, 53-63.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18(1), 10-30.
- Qualtrics Labs, Inc. (2017). Qualtrics [survey software]. Provo, Utah: Qualtrics.
- Rubin, D.L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511-531.
- Smith, L. (1983). English as an international language: No room for linguistic

- chauvinism. In L. Smith (Ed.), *Readings in English as an International Language* (pp. 7-11). New York: Pergamon.
- Smith, L. & Bisazza, J.A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning*, 32(2), 259-269.
- Smith, L. & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, 13(3), 371-380.
- Speelman, D., Spruyt, A., Impe, L., & Geeraerts, D. (2013). Language attitudes revisited: Auditory affective priming. *Journal of Pragmatics*, 52, 83-92.
- Teller, A. (2016). *Astro Teller: The unexpected benefit of celebrating failure* [video file]. Retrieved from [https://www.ted.com/talks/astro\\_teller\\_the\\_unexpected\\_benefit\\_of\\_celebrating\\_failure/transcript](https://www.ted.com/talks/astro_teller_the_unexpected_benefit_of_celebrating_failure/transcript)
- Tokumoto, M. & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes*, 30(3), 392-408.
- Verhagen, T., van den Hooff, B., & Meents, S. (2015). Toward a better use of the semantic differential in IS research: An integrative framework of suggested action. *Journal of the Association for Information Systems*, 16(2), 108-143.
- Walters, K. (2007). Language attitudes. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich and A. Zaborski (Eds.), *Encyclopedia of Arabic language and linguistics* (Vol. 2) (pp. 650-664). Leiden, The Netherlands: Koninklijke Brill NV.
- Zahn, C.J. & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4(2), 113-123.