Abstract

To assess the link between accent and intelligibility, previous research has heavily relied on human responses, demonstrating that human assessment is influenced by such subjective factors as linguistic proficiency (Pihko, 1997), familiarity (Smith, 1992), age (Burda, 2000), reactions to speakers’ kinetic cues (Bara, 1992; Rubin, 1992), and dialectal differences (Smith & Bisazza, 1982). In particular, language attitude studies have alerted us that nonstandard speakers are regularly regarded not only as suffering a pathological condition (Regional Rehabilitation Hospital, cited in Weinberger, 2000), but also as sounding less ‘pleasant’ and ‘correct’ than their standard counterparts (Scales et al., 2006). To separate these affective-attitudinal factors from research on accent-intelligibility correlation, this paper elaborates a phono-numerical measure on the grounds of linguistic evidence. Although listener judgments provide crucial clues for understanding how accent affects intelligibility, this measure aims to compare sound differences and explore their effect on phonological intelligibility as the linguistic index of perceptual intelligibility.

Keywords: Accent; Intelligibility; Phono-numerical analyses

1. Introduction

Accent has long been found to be a significant factor affecting intelligibility because listeners often experience difficulty in understanding utterances that differ from the sound patterns to which they are accustomed (Cheng, 1996; Gumperz, 1982, 1984, 1992; Jenkins, 2000; Kachi, 2004; Nishi, 2002; Rogerson-Revell, 2007). However, intelligibility has been found to be influenced by various subjective factors. Pihko (1997) observed that nonnative speakers’ listening comprehension of different English varieties correlates significantly with their general language proficiency. In addition, Smith’s (1992) findings reveal that the more familiar listeners (native and nonnative alike) are with other varieties of English, the better they can understand the speakers of such English varieties.

Furthermore, Burda (2000) noticed that older informants have greater difficulty understanding speakers with accents. Moreover, Bradlow and Pisoni (1999) found that nonnative listeners have more difficulty than native listeners with lexically difficult words, particularly at a fast speech rate. Additionally, Bara’s (1992) findings indicate that nonnative speakers who use more kinesic cues are usually rated by American English speakers to be more intelligible and comprehensible than those who do not use kinesics, although animated speakers may not be more linguistically competent. She explained that the kinesic
cues which affect perceived intelligibility and comprehensibility come primarily from facial expressions, the use of gaze, and general hand movements.

Most crucially, language attitudes towards speakers’ varying degrees of accentedness have also been found to interfere with mutual intelligibility. Those who have negative attitudes towards language variation tend to overstate the unintelligibility of nonstandard English speakers. Preston’s (1996) findings demonstrate that American southerners are often characterized by those from other states as incorrect and unintelligible, although they generally display affirmative attitudes towards Southern American English. Studies such as this reveal that a listener’s attitude towards other accents can affect perceptual comprehension, instead of reflecting actual intelligibility.

African-American English speakers have also been regarded by some Americans as unintelligible because they frequently pronounce words inaccurately, saying, for instance, aks for ask, tes for test, and cry for its third-person singular form cries (see the Oakland Ebonics controversy, Rickford, 1999). These sound deviations, in fact, are neither random nor pathological but, rather, reveal phonological patterns—metathesis, syllable-final consonant cluster simplification, and the absence of the third-person singular suffix, respectively (Green, 2002).

From her analysis of the group discussions which followed listening to a recorded segment of accented speech, Mettler (1989) found that some speakers of English tend to hold negative judgments of nonnative English utterances although they are able to understand what the nonnative speakers are saying. Again, this study indicates that even when an accent does not affect comprehension, negative attitudes might lead listeners to downgrade perceived intelligibility.

Although prejudice might penetrate listeners’ assessments of accented utterances, Munro et al. (2006) commented that participants “can choose to downgrade, ignore, or pay little attention to a L2 speaker” in the dictation and rating tasks (p. 19). When listeners constrain their subjective attitudes towards accented utterances, they are able to rate speakers’ comprehensibility on a dispassionate, if not objective, basis. In fact, most of the earlier studies presented above have shown that informants are capable of performing tasks when being asked to listen to L2 speech utterances and write down what they hear, or make judgments concerning comprehensibility and accentedness, as demonstrated from high inter-rater reliability drawn on a wide range of native and nonnative listeners. Participants in perception experiments are conscious that their performances are being assessed; therefore, they would try hard to complete the tasks.

Accordingly, to examine intelligibility with listener judgments, we need to screen our informants to recruit those aware of the sociolinguistic ecology of World Englishes espoused earnestly by such scholars as Kachru (1992, 1994, 2006), Jenkins (2000, 2003), Smith and Forman (1997), Wolfram (1991), Trudgill (2002, 2004), Rickford (1999), Seidlhofer (2004), Nickerson et al. (2005), and Kirkpatrick (2007). Otherwise, listeners’ perceptions of speech stimuli might not reflect speakers’ actual intelligibility (or relatively speaking, the respondents’ own listening skills) but, instead, reveal their own negative responses to accented speakers because of impatience and inexperience with nonstandard accents or others rather than theirs. This vigilance is necessary because, whenever we would like to conduct an evaluation or an assessment, we need to train our participants to adhere to the equitable principle explicitly given in our research. It is thus suggested that, to deconstruct the widespread preconception of nonstandard accents as fragmented and incomprehensible regardless of the context, future research needs
to train participants to deploy a well-articulated socio-phonological principle for a sound and fair evaluation of different accents before dictation and rating tasks.

In view of various parameters lurking in human responses to different accents in terms of intelligibility, Cheng (1996) asserts that an equitable measurement based on the comparison of sound systems is needed to address the relationships between accent and intelligibility. He remarks that prior research on “speech intelligibility was, by and large, the percentage of correct scores made by human subjects. While these tests are important for evaluating human understanding of speech, the results vary from subject to subject and therefore do not account for the mutual intelligibility of languages as systems” (my emphasis, p. 283). He proposes to make “a distinction between systemic mutual intelligibility, which describes a pair of languages in terms of their systems and correspondence patterns, and participant mutual intelligibility, which is based on […] personal experience and linguistic ability of individuals” (my emphasis, p. 280).

In contrast to perceptual intelligibility based on listeners’ opinions or judgments, this study elaborates a phono-numerical measure to explore phonological intelligibility on the grounds of the differences in sound systems. This approach separates subjective attributes from the analysis of intelligibility and focus only on the effect of the phonological distance between two distinct English speakers on their mutual intelligibility. This measure does not mean to negate the contributions of listener judgments in research on perceptual intelligibility. Rather, it aims to provide a phonological approach to accent-intelligibility correlation as a supplement to various existing participant-driven methods.

2. Literature Review

2.1 Definitions of Accent and Intelligibility

According to Crystal’s (1997) definition, accent refers to a speaker’s particular way of pronouncing words that often associates the speaker with those speaking the same native language and sharing similar sociolinguistic backgrounds. What then is intelligibility? Smith (1988; 1992) regarded it as the capability of word recognition. He further contrasted intelligibility with two other words: comprehensibility and interpretability. Comprehensibility denotes the understanding of the literal meaning, similar to Austin’s (1962) concept of locutionary force. By comparison, interpretability refers to the grasping of the intended meaning underlying a given utterances, equivalent to Austin’s (1962) illocutionary force. In brief, intelligibility refers to word identification; comprehensibility means the extent to which an utterance can be understood; interpretability denotes the understanding of the intention underlying the discourse.

By contrast, Munro, Derwing, and Morton (2006) defined intelligibility “as the extent to which a speaker’s utterance is actually understood and emphasized the importance of distinguishing this notion from comprehensibility, which refers to the listener’s estimation of difficulty in understanding an utterance” (p. 2). They elucidated the importance of the distinction between intelligibility and comprehensibility because “two utterances that are fully intelligible might entail perceptibly distinct degrees of processing difficulty, such that they are rated differently for comprehensibility” (p. 2). Therefore, in their numerous studies on L2 utterances, Munro, Derwing, and Morton examined the impact of accentedness on both intelligibility and comprehensibility by asking the listeners to rate L2 utterances for comprehensibility as well as accentedness after dictation tasks (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995, 1999; Munro, Derwing, & Morton, 2006).
2.2 Who Judges Intelligibility?

Considerable research on the intelligibility of English as a second/foreign language (ESL/EFL) centers on how nonnative speakers are intelligible to native speakers, disregarding the fact that interactions between nonnative speakers are rapidly increasing worldwide. Native English speakers have long been regarded as the only legitimate judges of English international intelligibility, but English is used extensively as a *lingua franca* in international communication between two nonnative English speakers without the presence of native speakers (Crystal, 2003; Nickerson et al. 2005; Rogerson-Revell, 2007).

The assumption that native speakers are the sole arbiters of international intelligibility in the use of English is what Kachru (1994) termed “the interlocutor myth” (p. 13). Kachru (1994) commented that English has often been taught across nations to interact with native English speakers, overlooking the widespread interactions between two nonnative speakers. Nelson (1992) remarked that “nonnative creativity, and indeed, diachronic changes of all sorts, are constantly judged in comparison to native models” (p. 336). Because of the growing population of nonnative English speakers in the world, Nelson (1992) proposed that English speakers, native and nonnative alike, should discard Anglo-speakers’ ethnocentric views on how English should be used and spoken in international interaction. As Munro et al. state (p. 114), “there is no a priori reason to assume that native listeners’ responses are representative of the reactions that might come from the target linguistic community; nor is there any reason to suggest that the evaluations of native listeners are inherently more valid or meaningful than those of any other group.”

In contrast to studies that only use native speakers as intelligibility judges, some scholars have proposed to invite both native and nonnative English speakers to examine each other’s intelligibility. Bansal’s (1969) study might be the first attempt to examine to what extent educated Indian speakers of English were intelligible to other educated nationalities; he asked both native and nonnative speakers to repeat or write down the recorded words or sentences. His findings show that educated Indian speakers were approximately 70% intelligible to educated people from other countries.

Moreover, Smith (1987) invited both native and nonnative speakers to rate each other’s varieties of English. The nonnative speech samples were recorded by speakers from China, India, Indonesia, Japan, Papua New Guinea, the Philippines, Taiwan, the UK, and the US. The listener groups included 10 Japanese learners of English, 10 native speakers of American English, and a mixed group consisting of one native and eight nonnative speakers, each from a different country. His findings indicate that the participants, native and nonnative alike, with greater exposure to other varieties of English perform better on intelligibility tasks than those without such exposure.

Earlier studies have indicated that shared L1 backgrounds and extensive exposure to a type of accent may lead to an advantage in hearing utterances from one’s own accent (Smith & Bisazza, 1982; Gass & Varonis, 1984; Smith, 1992; Jenkins, Modiano, & Seidlhofer, 2001). However, recent empirical studies have presented inconsistent findings. Major, Fitzmaurice, Bunta, and Balasubramanian (2002) observed

---

1 ESL refers to the acquisition of English as an associate language for intra-national communication, whereas EFL denotes the acquisition of English for interaction mainly with native English speakers. By contrast, English as an international language (EIL) emphasizes the use of English in interaction with not only native English speakers but also nonnative English speakers (Jenkins, 2000).
that the advantage in understanding accented utterances from speakers who share the same mother tongue is probably small and sporadically observable. They reported that their Spanish speakers showed a small intelligibility advantage when hearing Spanish-accented utterances; furthermore, the Chinese and Japanese speakers did not show any advantage in hearing their own accented English utterances.

It is noteworthy that Munro, Derwing, and Morton (2006) found that Cantonese, Japanese, Mandarin, and native English listeners responded similarly to the utterances of intermediate-level ESL speakers, regardless of whether they share linguistic backgrounds. In their investigation, the listeners from diverse speech communities responded correspondingly not only to distinct accents, but also to intelligibility and comprehensibility. Their Japanese and native English participants agreed quite strongly on the most and least intelligible of the 48 speakers. Moreover, the native English listeners understood the Japanese speakers as did the Cantonese and Mandarin listeners. They stated that, “if one student (or teacher) genuinely finds another student’s utterances difficult to understand, there is a likelihood that other students will have a similar experience” even though they speak the same native language (p. 18). Accordingly, they suggested that native and nonnative listeners’ judgments of L2 accentedness, intelligibility and comprehensibility might be fairly comparable. They concluded, “Although native English listeners’ assessments of foreign-accented English speech should not be viewed as inherently any more valid or meaningful than those of nonnative listeners, the findings of this study give us no reason to suggest that they are any less meaningful than those of the other groups either” (p. 16).

On closer inspection, although Munro, Derwing, and Morton (2006) found that native English listeners shared similar responses with EFL listeners to ESL utterances, whether both native English and ESL listeners might also respond similarly to ESL speech is left unanswered. In the Outer Circle like China, Japan, and Spain as classified by Kachru (1986), most EFL speakers usually do not use English in their interactions with their fellow nationals sharing the native languages, whereas most ESL speakers in the Expanding Circle such as India, Singapore, and Philippines, frequently communicate in English on a daily basis, particularly when they do not speak the same local languages. Because ESL speakers in a nation interact frequently in English with one another, it is presumable that ESL listeners might display better intelligibility and comprehensibility from their own accents than native listeners, particularly when the ESL utterances are rated as strong accentedness. This hypothesis merits inquiry and verification in future research.

A critical issue yet to be concerned is that familiarity with a given accent does not necessarily denote equivalent experiences with various sociolects associated with the accent. In the case of EFL, it appears that nonnative speakers have more exposure to the acrolectal English accents than mesolectal and basilectal ones because they learn English most of the time from their English teachers, presumably acrolectal speakers. Accordingly, intermediate-level EFL listeners might exhibit better intelligibility for acrolectal utterances than for basilectal and mesolectal EFL utterances from their own ethnic accents. It is also possible that acrolectal EFL speakers, particularly English pronunciation teachers, might better understand their basilectal EFL students than native English speakers because they are usually conscious about their basilectal students’ pronunciation difficulties that they have experienced themselves in the process of learning English. In this regard, future research might explore the effects of both speaker and listener EFL proficiency on intelligibility.

Most importantly, although substantial exposure might familiarize listeners with a given accent, it
does not necessarily enable them to become aware of the sound differences between the accent and their own, or to be more precise, the phonological differences underlying their respective accents. Scales et al. (2006) reported that their 11 Taiwanese listeners had highest accuracy rate in identifying the Taiwan English accent than American, British and Mexican accents, but they did not find it easiest to understand; rather, they found the American accent easiest to comprehend. This gap between experience and cognition has also been evidenced in research on the role of awareness in second language acquisition (Schmidt, 1992, 1995; Yang, 2004). Adult L2 learners tend to mishear the pronunciation of a new word, despite native speakers’ repetitive demonstrations. Without an adequate contrastive analysis of the L1 and L2 sound systems, adults often fail to effectively improve their L2 listening comprehension and speaking skills. The lack of linguistic knowledge might be why accent familiarity and experience exposure play a minor role in mutual intelligibility, unless they are accompanied by the knowledge of the accented English sound patterns in question.

Therefore, a growing number of scholars have emphasized that, to use English for international communication, we need to learn about the realistic usage of English by describing and teaching the varieties of English used by our frequent interlocutors. Spichtinger (2001) maintained that Europeans should learn from one another’s ways of appropriating English for their own European purposes. Seidhofer (2004) also stressed that corpus-based analyses of realistic English usage are needed to develop adequate teaching materials for the use of English as a lingua franca for international business communication. Moreover, Scales et al. (2006) recommended (p. 735):

*Instead of a single pronunciation model, English language learners could hear, analyze and compare key features among a variety of accents. Such an approach would address both intelligibility and listening comprehension, increasing communicative flexibility and respect for accent diversity.*

Likewise, Kirkpatrick (2007) proposed that English language teachers need to “understand how different varieties of English have developed linguistically and the ways in which they differ phonologically, lexically, grammatically, rhetorically and culturally” (p. 195). Obviously, much work on the teaching of English for international communication is in urgent need to recognize, respect, describe, and compare distinct varieties of English for effective and harmonious transnational interaction.

To sum up, in the use of English for international communication, it is inappropriate to regard native speakers as sole judges of intelligibility. Although some studies found native speakers’ responses similar to nonnative ones, one cannot ignore nonnative speakers as participants using and speaking English as a lingua franca for transnational interaction. There is a pressing need to uphold nonnative speakers’ call for national/ethnic identities and self-esteem. Accordingly, research on a variety of issues regarding accent, mutual comprehension, talk practices and communication styles needs to invite both native and nonnative speakers, beyond the Anglo-centric framework of native speakers as the representative judges of English international intelligibility.
3. Methodology

3.1 Earlier Approaches

Dictation tasks have been commonly adopted to examine verbal intelligibility (Bent & Bradlow, 2003; Brodkey, 1972; Burda et al., 2003; Derwing & Munro, 1997; Munro, Derwing, and Morton, 2006). Listeners are asked to use standard orthography to write out the utterances they hear; the number of the words they correctly transcribed is regarded as an index of speech intelligibility. In addition, Munro et al. (2006, p. 112-113) noted that other approaches have also been employed by other scholars to examine the relationships between utterances and intelligibility, including listening comprehension tests (Anderson-Hsieh & Koehler, 1988), cloze tests (Smith & Rafiqzad, 1979), responses with pictures to speech stimuli (Smith & Bisazza, 1982), recounting (Perlmutter, 1989), and truth value judgment (Munro & Derwing, 1995).

Remarkably, previous studies only employ a small number of speech stimuli for participants to identify words or utterances. Waters (2002) only used four minimal pairs (cap/cab, pick/pig, pot/pod, beet-bead) to test word-recognition intelligibility. He actually focuses merely on the pronunciations of CVC words. Because his findings indicate that American English listeners often cannot distinguish whether the word-final consonants of the test words spoken by nonnative speakers (in this case, Japanese and Taiwanese) are voiced or voiceless, he concluded that, to enhance intelligibility, spoken English instruction should include the exercises of producing and distinguishing words with final voiced/voiceless stops.

Additionally, Nishi (2002) only adopted a set of 270 words that included 15 vowels in the /hVd/ context to investigate the relationships between Japanese-accented English speakers’ interphonology and their intelligibility to Americans. Her results show that Japanese English learners’ intelligibility to Americans relies on the degree of the vowel deviation. Another problem with the above studies is that some of the test words are uncommon or culturally specific words that most nonnative speakers might not use in intercultural communication, such as pod, beet, hod, hawed, and hoed.

Moreover, Munro and Derwing’s (1995) only employed the excerpts of extemporaneous English produced by 10 Mandarin-accented English speakers to explore L2 intelligibility. Similarly, Scales et al. (2006) deployed a one-minute audio about insects to examine accent perceptions. Because earlier studies like these only investigate the perceptual intelligibility drawn on a small number of words and utterances, it remains unexplored how distinct accents affect general oral intelligibility. In addition, whether the listener is familiar with the content of the audio is another factor affecting the degree of intelligibility.

By contrast, Cheng (1996) illustrates how the mutual intelligibility of Chinese dialects might be calculated through a systematic comparison of phonological affinity. Reed and Spicer (1952) also designed correlation methods to quantify linguistic distance. Similarly, Y. Wang (1960) employed a lexicostatistical method to examine linguistic relationships among five Chinese dialects. W. Wang (1987) also adopted a quantitative measurement to determine linguistic affinity. Lu and Cheng (1987) further explored Chinese dialect affinity by comparing syllable differences.

These techniques have been developed by Chinese linguists since 1940s. However, they are not readily applicable to quantifying the phonological divergence of different English varieties because Chinese differs infamously from English in terms of their phonological systems. Of particular concern here is how we can quantify sound differences and measure their effects on systemic intelligibility. To this
end, this study aims to explore the following research question:

How can two distinct English accents be quantified to compute their phonological intelligibility?

To address this question, this paper elaborates a phono-numerical measure to which I now turn in the subsequent section.

3.2 A Phono-Numerical Measure

Although intelligibility has long been assessed by human subjects for different purposes, “phonological differences in the language systems under consideration,” Milliken and Milliken (1996) remark, “provide crucial keys to intelligibility unavailable through other approaches” (p. 15). Grimes (1989) also contends that, because greater phonological distance between two dialects always relates to lower mutual intelligibility, the communication networks of different dialects within the optimal framework of linguistic patterns merit inquiry and investigation. Accordingly, this study proposes a phono-numerical measure to analyze phonological intelligibility.

This measure employs a numerical calculation based on the similarity of the syllable onset, the nucleus, the coda, and the stress in a poly-syllabic word. Each syllable element in a word is given the same value because no evidence has demonstrated that one syllable feature is more salient than others. Although the nucleus (i.e., the vowel) is purportedly a prominent feature in a mono-syllabic word, syllable-initial and syllable-final consonants are equally crucial to distinguish a given word from others, as shown in minimal pairs, such as peak/beak and seat/seek.

In addition, this measure assigns the word stress the same value as other syllable elements because it is also uncertain whether it is more or less significant than other sound features. According to Bansal (1969), some Indian English speakers’ stress misplacements tend to lead other nationals to misidentify some test words as others, like character as director, written as retain, atmosphere as must fear, yesterday as or study, prefer as fearful, correct as carried, and about as come out. Nevertheless, he did not quantified misplaced stresses, nor did he examined whether their effects on word-recognition intelligibility are significant or not.

Notably, Jenkins (2000) observed that non-understanding among intermediate-level or mesolectal English speakers is rarely caused by differences in the word stress unless a different stress is used in combination with a segment substitution. Therefore, it is unclear whether the word stress is more crucial than other syllable segments in the identification of spoken English words. Without specific evidence of portraying the auditory scale for different syllable elements, this study considers each syllable element as having the same weight in the calculation of phonological intelligibility.

However, this measure does not consider intonation because it aims to investigate word-recognition intelligibility based on differences in word pronunciation. Prior research on real-world communication has indicated that discursive prosody does not influence intelligibility, although it may reveal speakers’ intentions underlying literal utterances. Numerous studies by Gumperz (1982, 1984, 1992) have demonstrated that intonation has much less influence on word recognition than on comprehensibility and interpretability. For instance, the expression How can I help you? with the default falling tone sounds ‘impersonal,’ ‘aloof,’ and ‘inapproachable,’ but a rising tone sounds ‘friendly,’ ‘warm,’ and ‘enthusiastic.’
As a result, native speakers may misunderstand nonnative speakers’ intentions when decoding their falling tones from the perspective of their own pragmatic usage manifested in intonation patterns. Taken together, as far as word-recognition intelligibility is concerned, Gumperz’ studies have shown that differences in intonation alone do not lead to non-understanding.

Deterding and Kirkpatrick (2006) also found that syllable-based rhythm does not affect ASEAN English speakers to talk with one another. Although such a speech rhythm may sound childish, exaggerative, annoying, or sarcastic to native listeners in various contexts (Crystal, 1995), the clear enunciation of each syllable in an utterance actually enhances mutual understanding to many participants from ASEAN regions (Kirkpatrick, 2004). Interestingly, Crystal (2003) remarks that the syllable-based style of rap chanting is becoming popular among young people.

In fact, Jenkins (2000) observed, from her database of Interlanguage Talk, that most intermediate-level or mesolectal English speakers are often unaware of, or are unable to produce, intonational cues, although intonation carries a wealth of pragmatic information to native speakers. She noticed that mesolectal English speakers largely decode each other’s messages based on word recognition—that is, the pronunciations of the intended words, rather than their intonation patterns. According to her survey, non-understanding between two mesolectal English speakers is caused mainly by segmental, not suprasegmental, differences. She reported that the misuse of syntactic contrastive stress rarely results in non-understanding without the combination of segmental errors. Therefore, she concluded that sentential prosody is irrelevant to mutual understanding.

In their recent paper entitled Understanding English Intonation, Clark and Wharton (2009) explained that intonation does not encode semantic information, but a non-default intonation might divert the hearer to the intended, albeit less accessible, interpretation, as shown in the following example:

(a) Billy insulted Tim and then he hit him.
(b) Billy insulted Tim and then hé hit him.

Simply put, pitch differences do not affect word identification and literal understanding but, rather, listeners’ perceptions of intended messages.

Discursive stress does not provide extra semantic meaning either. Although it reveals new information in the context, it does not affect mutual understanding. Deterding and Kirkpatrick (2006) reported from their conversation analyses that Singapore English speakers regularly use the heavy falling stress to end an utterance, but their interlocutors from other ASEAN nations do not find it to be a communication problem; rather, their dialogue went smoothly, as illustrated below (p. 401):

MMal: because there are a lot of students who are weak in English and they go to such schools just to learn ENGLISH.
FBrun: we have the government schools and the private SCHOOLS, but teaching of English begins in kindergarten.
FMyan: I love teaching and I enjoy TEACHING.
MThai: it was meant for only a h- a holiday a three-day HOLIDAY.
MCamb: erm English is very new and very few people speak ENGLISH.
Finally, this study does not consider the acoustic nuances of a vowel because they do not influence word recognition either (Nishi, 2002). For instance, the vowel /i/ appears in the sound inventories of both Mandarin and English, but Chen et al. (2001) found that the second formant (F2) of the vowel in Mandarin is lower than in American English; this means that Mandarin speakers produce the vowel /i/ with less advancement of the tongue as compared with American English speakers. This articulatory nuance, however, is not salient enough to be perceived to be another phoneme, such as /e/ and /α/. Mandarin learners of English actually often transfer Mandarin /i/ to English /i/, but native English speakers seem not to have a problem identifying such words as *tea*, *see*, and *seat* spoken by Mandarin speakers. Such a positive transfer is never described as an interphonological feature causing sound deviation from English (Yang, 2004; Chung, 2006). Therefore, a vowel transcribed by the International Phonetic Alphabet (IPA) as the phoneme for different languages like /i/ for the high, front vowel in both Mandarin and English is regarded as the same phoneme.

In sum, this phono-numeric measure considers syllable elements (onset, nucleus, and coda) and lexical stress in a poly-syllabic word, excluding intonation, discursive stress and acoustic nuances. Each word is given the value of one, and each of the sound components in a word is given the same value. Accordingly, a sound component has the value of 1/N, which refers to the number of total syllable components in the word. This condition is presented below:

<table>
<thead>
<tr>
<th>Condition (a)</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If a syllable element in a word spoken by speaker A is the same as its counterpart spoken by speaker B</td>
<td>1/N (N = the number of total syllable elements in a word)</td>
</tr>
</tbody>
</table>

This measure employs Heinle’s Newbury House Dictionary of American English as a point of reference not only because it is accessible on the Internet, but also because General American English is well known worldwide due to its use as a teaching model adopted by many countries and its effects on global technology and media. Take the word *bag* for example. This word includes three elements: the onset /b/, the nucleus /æ/ and the coda /g/. According to Condition (a), each element possesses the value of 1/3. If an element is pronounced differently by two speakers, it loses its value, namely 1/3. This condition is specified below:

<table>
<thead>
<tr>
<th>Condition (b)</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If a syllable element in a word spoken by speaker A is phonemically different from its counterpart (such as the vocalic difference between /æ, α/, /α, o/, /α, ɔ/, /ɛ, d/, /θ, t/, /w, v/, /w, wh/, /dʒ, ʒ/, /l, r/, and /p, p/))</td>
<td>0</td>
</tr>
</tbody>
</table>

---

If no sound difference in a word is found in two speakers’ pronunciations, the word preserves its value set as one. This condition is shown below:

<table>
<thead>
<tr>
<th>Condition (c)</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If a word spoken by speaker A is the same spoken by speaker B</td>
<td>1</td>
</tr>
</tbody>
</table>

If speaker A articulates the word *bag* as the same as speaker B, they can recognize the intended word from each other’s enunciations. Put simply, their mutual phonological intelligibility of the word is one. However, suppose speaker A reads *bag* as *[bæg]*, whereas speaker B pronounces it as *[bæɡ]*. The two speakers differ in the vowel. Therefore, the vowel is not given the value of 1/3. By contrast, the onset and the coda are the same in the speakers’ pronunciations, thus maintaining their values. As a result, the word value is the sum of the three segments: $1/3 + 0 + 1/3 = 1/3$.

This measure is proposed to compute 1000 high-frequency words to explore phonological intelligibility. Accordingly, two speakers’ phonological intelligibility can be formulated below:

\[
\text{PhonoI} = \frac{(W_1+W_2+W_3+\ldots+W_{1000})}{1000}
\]

Note:

PhonoI: phonological intelligibility

$W_1$: the value of the first high-frequency word

The following section will detail the retrieval of the high-frequency words and justifies why 1,000 words are sufficient to examine *general* phonological intelligibility.

This measure also addresses the problem of homophones in word recognition. If a word spoken by speaker A becomes another word spoken by speaker B, speaker A’s phonological intelligibility to speaker B is zero. This condition is displayed below:

<table>
<thead>
<tr>
<th>Condition (d)</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d. If a word spoken by speaker A becomes another word spoken by speaker B</td>
<td>0 (homophones)</td>
</tr>
</tbody>
</table>

When speaker A says *bag* as *bug*, then he is unintelligible to others who distinguish the minimal pair. Accordingly, speaker A’s phonological intelligibility of the word is zero.

This measure also considers the degree of phonological divergence between two phonetically similar phonemes. For instance, the difference between /æ/ and /a/ is smaller than that between /ʌ/ and /o/ because the phoneme /æ/ is phonologically closer to /a/. By comparison, the phoneme /ʌ/ is more divergent from /o/ because the former is an unrounded central vowel, whereas the later is a back, round vowel. The following vowel chart exhibits the phonological distances among the common English vowels:
Accordingly, two similar phonemes, albeit distinct, preserve half of the value, which is set below:

<table>
<thead>
<tr>
<th>Condition (e)</th>
<th>Value ('weight')</th>
</tr>
</thead>
<tbody>
<tr>
<td>e. If a syllable element in a word spoken by speaker A is phonetically different from its counterpart, such as the groups /ʌ, ə, œ, /i, ɪ, /u, ʊ, /e, ɛ, o, ɔ, ø, /e̞, ə̞, and /ə̞, a̞/</td>
<td>1/2N</td>
</tr>
</tbody>
</table>

Take the word *sun* for example. It comprises three sound elements: the onset /s/, the nucleus /ʌ/, and the coda /n/. Therefore, each component is assigned the value of 1/3. Suppose speaker A enunciates it as [sʌn], whereas speaker B reads it as [sɔn]. Their vowel difference leads to the loss of the vocalic value according to Condition (b), which is, however, blocked by Condition (d). It appears to be a phonetic variation. Therefore, it does not lose the vocalic value, but it only preserves half of the value. By contrast, other syllable elements maintain their assigned values because no sound difference occurs in the two speakers’ pronunciations. Consequently, the word value is the sum of the elements: 1/3 + 0 + 1/3.

In particular, this measure addresses addressing unidirectional inherent intelligibility (UII), which is usually disregarded in participant-driven research that only explores how a group of listeners perceive an accented speaker’s intelligibility rather than vice versa. Between two English speakers there frequently exists an asymmetry of intelligibility. If speaker A uses the contrastive vowels of /i/ and /ɪ/, while speaker B merges /ɪ/ into /i/, then speaker A may misrecognize speaker B’s pronunciation, for example, of the
word *pick* as *peak*. In real-world communication, any confusion of this nature is usually cleared up with the help of contextual cues or discursive information (as in the sentence *you can peak whichever one you like*). However, if the two words are played back in isolation to speaker A, speaker A will misperceive speaker B’s intended word *pick* as *peak* because no external information is available for speaker A to decode speaker B’s merger of the vowel /i/ into /i/. By contrast, speaker B is likely to identify speaker A’s intended word *pick* because he is able to perceive speaker A’s pronunciation of the word *pick* to be different from his own pronunciation of the word *peak*.

If a phonemic difference causes the misidentification of a minimal pair in which the test word is a high-frequency word and also the intended word, then the word is given the value of 1.0. This value condition was decided because there is evidence that, if one word is more common than the other in the case of homophones, it is usually recognized as the intended word (Bradlow, & Pisoni, 1999). The retrieval of high-frequency words from a spoken English corpus will be detailed in the next section.

Conversely, if a phonemic difference results in the misidentification of a minimal pair in which both the test word and the mistaken word are high-frequency words, then the test word is given the value of 0.5. This condition was set because both of the words may occur simultaneously to listeners from the psycholinguistic perspective. In other words, a listener may recognize either word 50% of the time when the two words are both high-frequency words. The following condition is formulated to UII:

<table>
<thead>
<tr>
<th>Condition (f)</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If one of the minimal pair spoken by speaker A does not exist for speaker B</td>
<td></td>
</tr>
<tr>
<td>f1. and is a high-frequency word but its counterpart is not</td>
<td>1</td>
</tr>
<tr>
<td>f2. and is a high-frequency word like its counterpart</td>
<td>1/2</td>
</tr>
</tbody>
</table>

A real case of unbalanced phonological intelligibility occurs in the merger of the consonant /θ/ into /t/ reported in ASEAN English lingua franca (Deterding & Kirkpatrick, 2006; Young, 2003). Suppose speaker A contains the contrastive consonants of /θ/ into /t/, while speaker B merges /θ/ into /t/. Then speaker A will misrecognize speaker B’s pronunciation of the word, for instance, *three* as *tree*. Because speaker B mispronounces the word, his UII is zero. By contrast, speaker A distinguishes the minimal pair; therefore, his UII to speaker B possesses some value. Because the word *tree* is not a high-frequency word (which will be discussed in the next section), speaker B is very likely to perceive the intended word. According to Condition (f1), speaker A’s UII maintains the value of 1. To sum up, their systemic mutual intelligibility (SMI) is the average of their UIIs, namely the value of (0+1)/2=1/2

Another nonstandard feature found in Estuary English by Deterding in 2005 is the substitution of /f/ for /θ/. Suppose speaker B says *three* as *free*. Because he mispronounces the word as another, his UII is zero. Nonetheless, speaker A does not undergo the sound change. Therefore, speaker B will perceive it as either *three* or *free* because both of the words are high-frequency words. According to Condition (f2), speaker A’s UII is 1/2. As a result, their SMI of the word *three* is (0+1/2)/2=1/4.

A quick comparison with earlier reports on the sound variations discussed above reveals that the phonological calculation of the word *three* reflects the degree of perceptual intelligibility. Deterding and Kirkpatrick (2006) noticed that the merger of /θ/ into /t/ does not cause any problem in conversations among ASEAN speakers, whereas the merger of /θ/ into /f/ is found to often bring forth difficulty for
listeners from China (Young, 2003) and Japan (Date, 2005). The following table summarizes the comparison:

Table 1. A comparison between perceptual intelligibility and phonological intelligibility

<table>
<thead>
<tr>
<th>Sound feature</th>
<th>Perceptual intelligibility</th>
<th>Phonological intelligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>merger of /θ/ into /t/</td>
<td>No communication problem among ASEAN speakers</td>
<td>Speaker A’s UII = 1</td>
</tr>
<tr>
<td>three → tree</td>
<td></td>
<td>Speaker B’s UII = 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMI=1/2</td>
</tr>
<tr>
<td>merger of /θ/ into /f/</td>
<td>difficulty for Chinese and Japanese listeners</td>
<td>Speaker A’s UII = 1/2</td>
</tr>
<tr>
<td>three → free</td>
<td></td>
<td>Speaker B’s UII = 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMI=1/4</td>
</tr>
</tbody>
</table>

Table 1 demonstrates that, like the previous reports on perceptual intelligibility, the computation of phonological intelligibility indicates that the merger of /θ/ into /t/ is more intelligible than that of /θ/ into /f/. Unlike perceptual intelligibility, however, phonological intelligibility can enable us to analyze each interlocutor’s UII, whereas the former only focuses on listeners’ perceptions and judgments. To sum up, the following table combines all of the conditions for the analysis of phonological intelligibility:

Table 2. The proposed phono-numerical measure

<table>
<thead>
<tr>
<th>Condition</th>
<th>Value (‘weight’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. If a syllable element in a word spoken by speaker A is the same as its</td>
<td>1/N</td>
</tr>
<tr>
<td>counterpart spoken by speaker B</td>
<td>(N = the number of</td>
</tr>
<tr>
<td></td>
<td>total syllable elements</td>
</tr>
<tr>
<td></td>
<td>in a word)</td>
</tr>
<tr>
<td>If a syllable element in a word spoken by speaker A is phonemically different from its counterpart (such as the vocalic difference between /æ, α/, /α, ɔ /, /α, ɔ/, /ʊ, d/, /θ, t/, /w, v/, /w, wh/, /dʒ, z/, /l, r/, and /p, p/)</td>
<td>0</td>
</tr>
<tr>
<td>c. If a word spoken by speaker A is the same spoken by speaker B</td>
<td>1</td>
</tr>
<tr>
<td>d. If a word spoken by speaker A becomes another word spoken by speaker B</td>
<td>0 (homophones)</td>
</tr>
<tr>
<td>e. If a syllable element in a word spoken by speaker A is phonetically different from its counterpart, such as the groups /ʌ, α, ɔ, θ/, /i, ɪ/, /u, ʊ/, /e, ɛ/, /o, ɔ, ɔ/, /ʃ, z/, /l, r/, and /ʃ, θ, θ/.</td>
<td>1/2N</td>
</tr>
<tr>
<td>f. If one of the minimal pair spoken by speaker A does not exist in speaker B</td>
<td></td>
</tr>
<tr>
<td>f1. which is a high-frequency word but its counterpart is not</td>
<td>1.0</td>
</tr>
<tr>
<td>f2. which is a high-frequency word like its counterpart</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Note: The application order of the conditions is c > d > f > e > a > b.