The intelligibility cocktail: An interaction between speaker and listener ingredients

BETH ZIELINSKI – La Trobe University

ABSTRACT

The findings of previous studies investigating the relationship between speech production and intelligibility in speakers of English as a Second Language (ESL) are far from conclusive. The aim of the research reported in this article was to gain further insight into this relationship and thus shed light on the best ways to improve intelligibility in speakers who have this as their goal. Because intelligibility involves both the speaker and the listener, this article presents findings of a case study that explored the role of each in reducing intelligibility in a Vietnamese speaker for three native (Australian) listeners. In this article I propose that reduced intelligibility (RI) was the result of the interaction between listeners' processing strategies and a complex mix of non-standard features in the speech signal. The listeners appeared to rely on the syllable stress patterns and segments in the speech signal to identify the speaker's intended words, but for this particular Vietnamese speaker, non-standard segments seemed to play a greater role in reducing intelligibility than did non-standard syllable stress patterns.

Introduction

Improvement in speech intelligibility is an important goal for many learners of ESL because speech produced in a way that affects intelligibility has a serious effect on their ability to communicate in English. Morley (1994: 67) asserted that 'speakers with poor intelligibility have long-range difficulties in developing into confident and effective oral communicators', and an inability to communicate effectively can affect many areas of a person's life. Fraser (2000) argued that if non-native English speakers settling in Australia are unable to communicate effectively in spoken English, they might be at risk educationally, occupationally, professionally and socially.

To be able to improve second language (L2) speech intelligibility, however, we need to understand the ways it is influenced by the different non-standard features of speech production[1] commonly present in the speech of L2 speakers. The relationship between non-standard features of speech production and intelligibility is therefore the topic of much discussion and debate in the literature, and has become an important focus of L2 pronunciation research. In fact, Field (2005: 399) argued that 'the most pressing issue in L2 pronunciation research today is the quest to identify the factors that most contribute to speaker intelligibility'. Despite the discussion, and focus on intelligibility in the literature, however, empirical studies that investigate the relationship between non-standard features of speech production and intelligibility in speakers of ESL are few, and their findings are far from conclusive. Thus, the aim of the research reported in this article was to investigate this relationship.

Intelligibility

DEFINITION

Interpretation and comparison of previous research findings related to intelligibility in L2 speakers is sometimes difficult because of the many different terms used by different authors (see Jenkins 2000 for a discussion of the different terms used in the literature). It is therefore important that I clarify my definition of intelligibility before continuing. I view intelligibility as involving both the speaker and the listener, and use the term to describe the listener's ability to identify the speaker's intended words. Intelligibility is therefore defined as the extent to which the speech signal produced by the speaker can be identified by the listener as the words the speaker intended to produce. This definition is similar to that used by Field (2005: 401): 'the extent to which the acoustic-phonetic content of the message is recognisable by a listener', and to the term 'phonological intelligibility' used by Jenkins (2002: 86).[2]

Intelligibility involves both the speaker and the listener. The contributions of both the speaker and the listener therefore considered in the following review of the literature related to the influence of non-standard phonological features of speech production on intelligibility.

The speaker's contribution: Non-standard phonological features in the speech signal

A common thread in the literature is the relative impact of non-standard suprasegmental features (stress, rhythm and intonation) and non-standard segments (consonants and vowels) on intelligibility. In a quest to identify pedagogical priorities for improving intelligibility, researchers have often set out to determine whether one is more important than the other. However, although there is support in the literature that both non-standard suprasegmental and non-standard segmental features have the potential to influence intelligibility, there is, as yet, a lack of empirical evidence to support that one has a greater influence on intelligibility than the other.

Only a handful of studies have investigated the influence of non-standard phonological features on intelligibility as defined above.[3] Munro and Derwing (1995) and Derwing and Munro (1997) considered the relationship between the intelligibility of connected speech and broad measures of both suprasegmental and segmental features of L2 speech production. In both studies, the researchers had native speakers of Canadian English orthographically transcribe utterances selected from the speakers' descriptions of the story told by a series of cartoons. The speakers were from a range of native language backgrounds, including Mandarin (Munro and Derwing 1995), Cantonese, Japanese, Polish and Spanish (Derwing and Munro 1997). Neither study found a strong relationship between intelligibility scores (calculated on the basis of the number of words transcribed accurately by the listeners) and non-standard suprasegmental features: neither judgments of the nativeness of intonation (Munro and Derwing 1995) nor judgments of goodness of prosody (Derwing and Munro 1997) correlated significantly with intelligibility scores for the majority of listeners. Likewise, neither study found a strong relationship between intelligibility scores and non-standard segmental features; the number of phonemic errors (defined as the deletion, insertion or substitution of a segment) failed to correlate significantly with intelligibility scores for the majority of listeners (although they did find stronger relationships between comprehensibility judgments and non-standard aspects of L2 speech – see discussion below).

Bent, Bradlow and Smith (in press) also used listener transcriptions to measure intelligibility. They investigated the effect of non-standard segments in different word positions on the intelligibility of sentences read aloud by L2 speakers from a Mandarin first language (L1) background. Each speaker was given an intelligibility score (across all sentences they read), calculated on the basis of the percentage of key words correctly transcribed by five native (American) listeners. The intelligibility scores were then related to segments produced correctly in the sentences. Bent, Bradlow and Smith found a significant correlation between intelligibility scores and segment production accuracy scores (speakers with higher segment production scores tended to have higher intelligibility scores). For some speakers, however, a lower intelligibility score did not mean that they had a low segment production accuracy score, and the speaker with the lowest intelligibility score had produced 83% of her segments accurately. The authors surmised that other non-standard features, not included in the analysis (eg word-level syllable stress patterns), might have contributed to her low intelligibility score. Another finding of the study was that accurate vowel production and accurate word initial consonant production were significantly correlated with intelligibility scores, whereas overall consonant accuracy and accurate consonant production in other word positions were not.

The findings of the above studies may have been influenced by the methodology used by the researchers. Intelligibility scores based on words (or key words) transcribed accurately give no information about where intelligibility is reduced in an utterance. We know from the literature that listeners draw on knowledge including context and syntactic and lexical knowledge to identify words in connected speech (see, for example, Bard, Shillcock and Altmann 1988; Wright, Frisch and Pisoni 1996–1997; Cutler and Clifton 1999; Moore 2003). Consequently, the listeners in the above studies may have found some words difficult to identify but were able to transcribe them accurately because of the context in which they were produced. Munro and Derwing (1995) and Derwing and Munro (1997) surmised that this might have been the case in their studies, and argued that an utterance might be accurately transcribed even though it is judged as difficult to understand. In both studies, as well as orthographically transcribing the speakers' utterances, the listeners rated each utterance for its 'comprehensibility' on a scale from 1 to 9, where 1 corresponded to 'extremely easy to understand' and 9 to 'impossible to understand' (see Munro and Derwing 1995: 79; Derwing and Munro 1997: 5). Comprehensibility was found to correlate significantly with intelligibility in both studies (for 83% of the listeners in Munro and Derwing 1995 and all of the listeners in Derwing and Munro 1997), but it seems that there was a stronger relationship between comprehensibility and suprasegmental and segmental features of L2 speech production than there was for intelligibility. Munro and Derwing found that, whereas comprehensibility correlated with judgments of the nativeness of intonation for the majority of listeners, intelligibility did not, and although comprehensibility correlated with phonemic errors for 44% of listeners, intelligibility did so for only 28%. Furthermore, although Derwing and Munro found that comprehensibility did not significantly correlate with goodness of prosody or the number of phonemic errors for the majority of listeners, the number of listeners for which it was significantly correlated with goodness of prosody was over four times as many as was the case for intelligibility, and with the number of phonemic errors, twice as many. It seems from these findings that non-standard features of L2 speech production may be more closely related to how hard something is to understand than to how accurately it can be transcribed, and listeners may have eventually been able to transcribe an utterance accurately, even though they found it difficult to identify the speaker's intended words.

Thus, calculating an intelligibility score based on words (or key words) transcribed accurately does not provide any information about which words were difficult to identify and which words were not, and therefore it is difficult to determine which non-standard features of speech production were related to reducing intelligibility and which were of no consequence. Although the above findings related to judgments of comprehensibility suggest that this might have been the case in the Derwing/Munro studies, these judgments did not assist in identifying the parts of the utterances where intelligibility was reduced. An added complication in the use of intelligibility scores based on words (or key words) transcribed accurately is that, when transcribing utterances, listeners may have transcribed some words inaccurately for reasons unrelated to the speaker's speech production (eg distraction by an earlier non-standard production, memory difficulties, spelling difficulties, unexpected word use).

Benrabah (1997) and Suenobu, Kanzaki and Yamane (1992) took a different approach. Although they used listener transcriptions to measure the intelligibility of utterances selected from connected L2 speech, rather than calculating intelligibility scores based on words transcribed correctly, they looked at the correspondence between non-standard features and transcription inaccuracies. Benrabah provided examples from connected speech produced by Indian, Nigerian and Algerian speakers, where words produced with non-standard stress had been transcribed inaccurately by the listeners. He observed that in their transcription of these words, the listeners relied heavily on the stress pattern of the word and totally disregarded the segmental information. Suenobu, Kanzaki and Yamane found that both non-standard segments and non-standard suprasegmental features influenced intelligibility in connected speech produced by Japanese speakers of English. When the listeners were listening to utterances that contained words produced in a non-standard way (target words), consonant deletion had the greatest effect on the intelligibility of the target words, followed by word stress errors, wrong pause insertion, vowel substitution, consonant substitution and vowel addition.

Although the above studies indicate that both non-standard suprasegmental features and non-standard segments may have an influence on the intelligibility of connected speech or sentences read aloud, we are still far from understanding the relationship between features of speech production and intelligibility. Some of the above studies focused on either suprasegmental features or segmental features, and some attempted to determine which had the greatest influence. There has been little investigation of the way in which non-standard suprasegmental and segmental features might combine or interact to influence intelligibility. For example, some of the inaccurate vowel productions reported by Bent, Bradlow and Smith (in press) as significantly correlated with intelligibility may in fact have been vowel changes related to the production of a non-standard syllable stress pattern. Also, none of the above studies provided insight into why the different non-standard features had an impact on intelligibility. In order to gain this insight, it is important to understand what features the listeners may have attended to in the speech signal in these studies, and why the different non-standard features may have misled them in their attempts to identify the speakers' intended words. As Field (2005: 400) argues, 'the psycholinguistics of first language (L1) speech processing provides an important key to an understanding of the factors contributing to intelligibility'.

The listener's contribution: The features of speech production important to native English listeners

All of the studies reviewed above had native English speakers as their listeners and, as the listeners in the research reported in this article were also native speakers of English, the phonological features of speech production that are important to native listeners are relevant here.[4]

Although the processes by which native English listeners derive a sequence of words from a stream of connected speech are not fully understood, previous research has indicated that they draw on a complex mix of suprasegmental and segmental features in the speech signal that are tailored to their native English phonology to do so (see for example Cutler and Clifton 1999). The rhythmic properties of the speech signal are important to them, as they draw on these properties to divide continuous speech into individual words, and to recognise what the individual words are. Cutler and Butterfield (1992: 218) described the speech rhythm of English as having a characteristic pattern of strong versus weak syllables, where strong syllables were described as those that 'bear primary or secondary stress and contain full vowels', and weak syllables as those that are 'unstressed and contain short, central vowels such as schwa'. Cutler and Butterfield found that when the speech signal was not clear, native English listeners depended on these rhythmic properties of the signal to identify word boundaries and tended to insert word boundaries before strong syllables and delete word boundaries before weak syllables. They argued that this reliance was due to the listeners' experiences with the English language, as the majority of words in English begin with strong syllables. Grosjean and Gee (1987) also argued that the rhythmic properties of the speech signal are important for listeners, and hypothesised that listeners use strong syllables to initiate a search for words in their minds, and that 'the stress pattern of a word, and the phonetics of its stressed syllable, are valuable information for the lexical access system' (Grosjean and Gee 1987: 149). Bond and Small (Bond and Small: 1983) support these ideas in their findings that listeners found changes in syllable strength with the associated vowel quality change disruptive, and found it hard to identify two-syllable words in which the position of the strongest syllable had been changed.

Strong syllables, and the vowels they contain, have been found by some researchers to be sources of important and reliable information for listeners. Bond (2003), in her research on listeners' misperceptions of speakers' intended messages (known as slips of the ear), found that vowels in strong syllables were very rarely misperceived by listeners, and therefore saw them as a source of reliable information. Also, Bond and Small (1983) found that listeners found changes to vowels in strong syllables disruptive, and had difficulty identifying two-syllable words in which the vowel in the strong syllable had been changed. Consonants in the speech signal are also seen as important to listeners. Stevens (2002), for example, argued that the way a consonant is produced depends on its position in the syllable. The way a consonant is produced can therefore help the listener determine word boundaries (because word initial consonants are also syllable initial). Consonants also contribute to the identification of words. Bond and Small (1983) reported that listeners found changes to consonants disruptive, and had difficulty identifying two-syllable words in which the voicing of syllable initial consonants had been changed. However, Bond and Small also found that changes to consonants were less disruptive to listeners than changes to vowels in strong syllables or the position of the strongest syllable.

The above studies suggest that English listeners, when relying on listening strategies tailored to their native English phonology, draw on both suprasegmental and segmental features in the speech signal; consonants, vowels (particularly those in strong syllables), and the rhythmic properties of the speech signal are all important to English listeners in the process of identifying a speaker's intended words. Thus, it would seem that non-standard production affecting these features is likely to mislead listeners and result in them identifying words other than those intended by the speaker. As noted above, previous studies have found that a range of different non-standard features influence intelligibility. The aim of the research reported in this article was to gain a better understanding of how non-standard features in the speech signal interact with the listener's processing strategies to reduce intelligibility when the listener is a native speaker of English and the speaker is an L2 speaker of English whose first language is Vietnamese. It was designed to explore (as described by Liss et al 1998: 2457) 'the interface between the speech signal and the listener's response to that signal' with the following questions in mind:

1 What features of the speech signal do the listeners rely on in their attempts to identify the speaker's intended words?

2 What non-standard features of speech production are implicated in misleading the listeners (and thus reducing intelligibility), and to what extent are they implicated?

Methodology

The findings presented relate to a case study involving four participants, a speaker and three listeners, and are taken from a larger research project, which also involved two other L2 speakers, one from a Korean and one from a Mandarin L1 background. The speaker was a 31-year-old male from a Vietnamese L1 background (from Hanoi) who had been in Australia for one year and two months at the time of the data collection. He was studying for a postgraduate qualification (Masters by research) at an Australian university and was a proficient speaker of English (reported IELTS [International English Language Testing System] score of 7.0 and TOEFL [Test of English as a Foreign Language] score of 580). The listeners were three female native speakers of (Australian) English, aged between 40 and 47 years of age. They all had tertiary qualifications in areas unrelated to Applied Linguistics or Teaching English to Speakers of Other Languages (building, primary teaching and food technology), reported no hearing difficulties, and had no particular experience with the Vietnamese language or with listening to Vietnamese speakers of English.[5]

Sixty-eight utterances selected from a recording of the speaker explaining the education system in Vietnam were presented in a different random order to each listener (see Appendix for examples of utterances). The listeners were told the general topic of the conversation from which the utterances were selected, and asked to orthographically transcribe each of the utterances: they were instructed to write down the words they heard the speaker say.

Given that the speaker's first language was Vietnamese, it is relevant to consider briefly the phonology of Vietnamese, which is different in many ways to the phonology of English. Vietnamese is a tonal language in which the majority of words have one syllable (Hwa-Froelich, Hodson and Edwards 2002). In connected speech produced by speakers from Hanoi, the majority of syllables are produced with the same level of stress, with alternate syllables slightly louder, and at least one per pause group produced with heavy stress (Thompson 1987). Thompson lists 19 different consonants and 15 different vowels used by Vietnamese speakers from Hanoi, some of which are similar to English consonants and vowels, and some of which are different. However, although all but one of the consonants can occur in the initial position of a syllable, only eight can occur in the final position, six of which are equivalent to the English consonants: /p, t, k, m, n, ŋ/. Consonant clusters never occur in the final position of syllables, and syllable final consonants are generally unreleased or weakly released (Hansen 2004; Thompson 1987). In the initial position of syllables, consonant clusters always involve /w/ in combination with non-labial consonants. Previous studies have found that when speaking English, Vietnamese speakers may have difficulty with syllable stress, vowels and consonants, particularly consonant clusters and consonants in the word final position (see for example Nguyen and Ingram 2004; 2005).

Identifying sites of reduced intelligibility

Rather than calculating intelligibility scores based on the number of words or key words correctly transcribed, I identified the parts of the utterances where intelligibility was reduced, that is, where any one of the listeners was unable to, or experienced difficulty in being able to, accurately identify the word/s the speaker intended to say. By identifying sites of reduced intelligibility (RI) in this way, the research effectively explored both intelligibility and comprehensibility as defined in the studies by Derwing and Munro discussed above.

To identify sites of RI, and gain some insight into how the listeners went about the task of identifying the speaker's intended words, I observed one listener at a time and used data from a number of different sources: the listeners' transcriptions (words transcribed inaccurately, words left out), their comments (eg *It sounds like 'walking' but it clearly isn't* or *Because he's talked about agriculture I can pick up that the word should be 'plants'*), and their behaviour while transcribing (eg facial expressions, sighs of exasperation, hesitation). When locating a site of RI, I looked for evidence from more than one data source to support my decision by cross-validation.

When using listener behaviour as evidence, I always checked for the reason for the behaviour, either by asking directly or from another data source. I was very careful not to be directive when asking the listeners about their responses, and mostly used open-ended questions.

The listeners were able to listen to an utterance a second time upon request, which meant that there were often two transcriptions for each utterance. A listener's first transcription, and related behaviour and comments, was my starting point in identifying sites of RI. Thus, if a listener transcribed part of an utterance inaccurately in the first transcription, and corrected it in the second, the first transcription was used in the analysis. The way the listener transcribed the utterance after hearing it the second time (second transcription), together with related comments, was a source of additional evidence used to clarify or add to the evidence gathered from the first transcription. One of the listeners listened to every utterance a second time, just to make sure, even when she said she was certain she had identified the speaker's intended words accurately. The other two listeners listened to most of the utterances a second time (43 and 46 utterances respectively).

Where there was a second transcription, a comparison of the two was interesting for a number of reasons. First, to clarify comments made about the first transcription, particularly those related to memory. Listeners sometimes commented that they had identified the speaker's intended words, but could not remember them to write them down. In such cases, however, there was no guarantee that the difficulty in remembering the words was not complicated by RI at that point in the utterance. If the listener was able to accurately transcribe the missing words in the second transcription, it is likely that it had been an issue of memory. However, if she was unable to transcribe the 'forgotten' words the second time, there may have been factors other than memory interfering with the ability to transcribe them. Second, the comparison between first and second transcriptions was one of the indicators that a listener may have relied heavily on context rather than phonological features to identify a word in the first transcription, thus suggesting that there was RI. Sometimes a listener transcribed a word correctly in the first transcription but changed it to an incorrect word in the second transcription. Usually the listener commented that the word sounded different the second time and no longer sounded like the word she thought she had heard the first time. Also, sometimes when the listener left a space for part of an utterance in the first transcription and indicated that she had no idea what the speaker had said, she was able to give an indication of what the word/s sounded like in the second transcription.

Data analysis

In order to capture the phonological features of speech that are important to English listeners (discussed above), I analysed the speech produced at each site in terms of three different levels of the phonological hierarchy: pause group, syllable and segment. Analysis at the pause group and syllable levels relates to the rhythmic properties of the speech signal, and analysis at the segment level, to the segments the speaker produced. At the pause group level, I used the speaker's placement of pauses to identify the boundaries between the smaller sections of speech within each utterance; at the syllable level, I judged the relative strength of each syllable in a pause group as the strongest in the pause group (**S**), strong but not the strongest (**s**) or weak (**w**); and at the segment level, the segments in each syllable were identified and phonetically transcribed. The graphic representation of the speech signal provided by ACID Pro 5 software was used as a crude confirmation of my judgments at the pause group and syllable levels. This graphic representation shows amplitude across time. At the pause group level, I only identified pauses that were confirmed on the graphic representation as zero amplitude. At the syllable level, the strongest syllables and the other strong syllables in the pause group were confirmed by referring to the amplitude at the relevant parts of the speech signal. On those occasions where my judgments were not confirmed by the graphic representations, I employed a second listener, experienced in analysing spoken English, to analyse the speech signal independently. Whichever judgment this listener agreed with (mine or the graphic representation) was used in the analysis. The segments were transcribed by a trained phonetician; where I disagreed with his transcription, I employed a second listener, experienced in transcribing spoken English, to transcribe the segments independently. Again, whichever transcription the listener agreed with (mine or the phonetician's) was used in the analysis.

Identifying non-standard features implicated in misleading listeners

Any pause group containing a section of speech with RI (identified as detailed above) was considered to be a site of RI. To identify non-standard features implicated in RI at these sites, I used the approach taken by Bond (1999: 61), who asserted that 'the most revealing way of describing complex misperceptions is to consider how the phonological shape of the misperception matched or failed to match the target utterance'. The phonological shape of each listener's response or attempt to identify words was compared to that of the speaker's production of those words. If the match between the two involved a non-standard feature for at least one listener, I identified that non-standard feature as implicated in misleading listeners. Non-standard features implicated at each of the three different levels of the phonological hierarchy (pause group, syllable and segment) were identified. The speaker's intonation pattern for each utterance was not included in this analysis and therefore the influence of a non-standard intonation pattern on intelligibility was not considered.

At 88 of the 109 sites of RI identified (80.7%) I was able to identify non-standard features of speech production that were implicated in misleading any one of the listeners, and these are the sites included in the analysis presented in the following section. In the 21 remaining sites I was unable to identify these features because the listener/s left a space, with no indication of what the speaker had said, or the phonological shape of the listener's response bore no resemblance to the speech produced. The majority of the 88 sites included in the analysis were sites of RI for more than one listener (n=65; 73.9%). Those that were sites for only one listener were also included in the analysis, however, as these gave further insight into how the different non-standard features misled those listeners. In total there were 169 listener attempts to identify the speaker's intended words that gave me an indication as to what non-standard features had misled the listeners at these 88 sites. At some sites, there were two or three listener attempts that provided this information, while at others there was only one. The non-standard features implicated in misleading listeners at any one site were a combination of features that misled each listener who made an attempt to identify the words at that site.

As noted above, sites of RI included listener responses where the listener was unable to identify the speaker's intended words, and responses where the listener was able to identify them but had difficulty in doing so. In the majority of listener responses the listener was unable to identify the speaker's intended words. Of a possible 264 listener responses (3 per site), 181 (68.6%) did not identify the words correctly, and 21 (8.0%) identified the words correctly but had difficulty in doing so. In two responses, the listener could not identify some words at the site and had difficulty identifying others, and in the remaining 60 (22.7%) responses the listener identified the words correctly.

The following example illustrates the difference that was often seen in listener responses at the same site of RI. At this site in utterance 3, Listener 1 could identify the word level with no apparent difficulty, Listener 2 identified it but found it difficult and drew on context to do so, and Listener 3 was unable to identify it. (The speaker had produced the word *level* as /levən/ with a **S w** syllable stress pattern.)

At that level
levən
S w

Listener 1: Transcribed *at that level* and didn't want to hear the utterance again.

Listener 2: Transcribed *at that level* but commented:

Now I'm not 100% sure that I understood ***level*** *to be level but by the time he got to the end, level sounded like* ***level****.*

She wanted to hear the utterance again, and after hearing it a second time commented:

He probably actually says ***at that leven*** *as opposed to* ***at that level****. I hear* ***at that level*** *because I know that that's what the word should be. If he just stood there and said* ***level*** *I would hear it as* ***leven*** *and I wouldn't know what he meant.*

Listener 3: Left a space for all three words and commented:

No idea what the start was.

She wanted to hear the utterance again, and after hearing it a second time was still unable to identify the words.

From this evidence, the non-standard feature identified as implicated in misleading listeners at this site was the production of /n/ rather than /l/ in the word final position.

Findings

In presenting these findings, I look at the interaction between the listening strategies used by the listeners and the non-standard features produced by the speaker, and thus consider the contribution of the listener and the speaker to the reduction in intelligibility. As will be shown, the listeners relied on speech processing strategies that drew on familiar suprasegmental and segmental features in the speech signal to identify the speaker's intended words, whether the speaker had produced them in a standard or non-standard way. A complex mix of non-standard suprasegmental and segmental features in the speech signal therefore misled them, thus contributing to reducing intelligibility.

The rhythmic properties of the speech signal

THE FEATURES THE LISTENERS RELIED ON

The listeners appeared to rely heavily on the speaker's syllable stress pattern when attempting to identify his intended words at sites of RI. In 87.6%

(n=148) of the 169 listener attempts to identify words at sites of RI, the listener replicated both the number and the strength of the syllables in the syllable stress pattern. For example, if a word or string of words produced by the speaker consisted of four syllables with a **s w S w** syllable stress pattern, the listener would replicate the number and strength of the syllables by identifying a word or words that also had four syllables with the first and third syllable produced as strong and the second and fourth produced as weak. Because they relied so heavily on the speaker's syllable stress pattern, a non-standard syllable stress pattern was implicated in misleading them at a number of sites of RI.

THE NON-STANDARD FEATURES IN THE SPEAKER'S PRODUCTION

At almost one third of the sites (28 of 88; 31.8%) the listeners replicated a non-standard syllable stress pattern produced by the speaker, and were thus misled into wrongly identifying the speaker's intended words. These included non-standard syllable stress patterns of three different types: those with non-standard syllable strength (eg *economics* produced with a **w S w s** pattern), non-standard additional syllables (eg ***have*** produced with /ə/ added at the end, resulting in a two-syllable word with a **S w** pattern) or the non-standard deletion of syllables (eg five and above, which should be produced as with a **S w w s** pattern, produced as five above with a **S w s** pattern). The most common type was the pattern with non-standard syllable strength (22 of the 28 or 78.6%), while the non-standard addition or deletion of a syllable was implicated at only four and two sites respectively.

When listeners were misled by syllable stress patterns of whatever type, they were also misled by accompanying vowel changes (the methodology in this study does not allow the separation of the two). If a syllable was produced as strong when it should have been weak, the vowel in that syllable was produced as full rather than reduced, and if a syllable was produced as weak when it should have been strong, the vowel in that syllable was produced as reduced rather than full. When a syllable was added, a vowel was also added, and when a syllable was deleted, a vowel was also deleted.

Thus, the listeners appeared to depend heavily on the speaker's syllable stress pattern at sites of RI and were misled when it had the wrong number of syllables or had syllables produced with non-standard strength. However, the syllable stress pattern was non-standard at less than one third of the sites and therefore provided accurate information at over two thirds. At these sites, non-standard segments played an important role in misleading listeners.

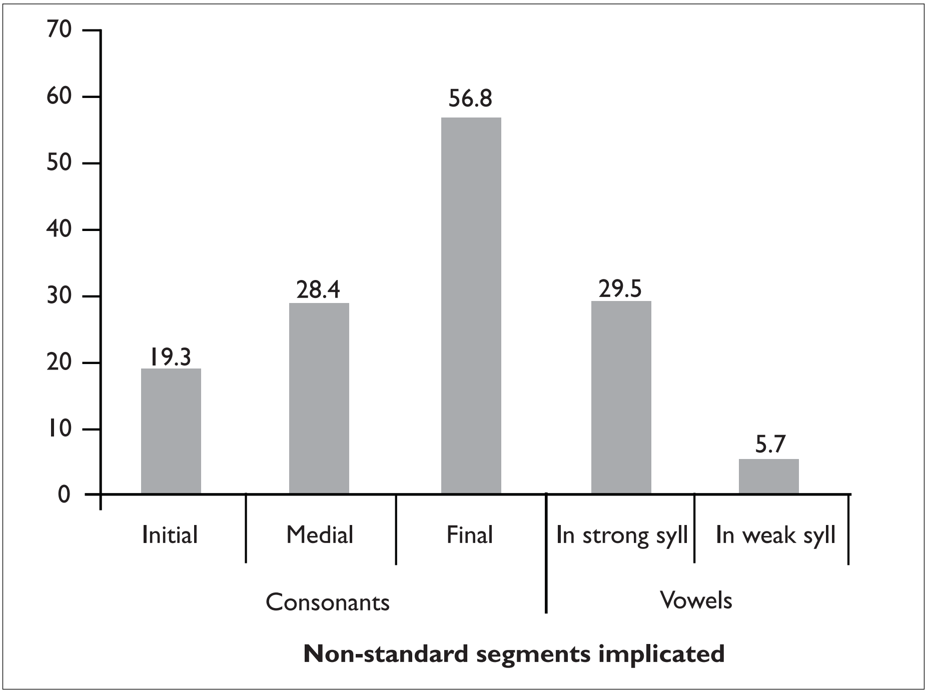

Figure 1: The percentage of sites of RI (n=88) at which non-standard consonants (initial, medial and final word positions) and non-standard vowels (in standard strong and weak syllables) were implicated in misleading listeners

Note. The total percentage of the categories is greater than 100% because at some sites more than one category was implicated.

Segments in the speech signal

THE FEATURES THE LISTENERS RELIED ON

The listeners appeared to rely on segments in the speech signal at sites of RI in their attempts to identify the speaker's intended words, whether they were standard or non-standard, as they were replicated to some extent in all but one listener attempt. However, unlike the case of syllable stress pattern, segments were not replicated exactly at the majority of sites: the listeners replicated the segments produced by the speaker exactly in only 37.3% (n=63) of their 169 attempts.

THE NON-STANDARD FEATURES IN THE SPEAKER'S PRODUCTION

Non-standard segments occurred at 80 of the 88 (90.9%) sites of RI, and were therefore implicated in misleading listeners more frequently than were non-standard syllable stress patterns. Both non-standard consonants and non-standard vowels[6] were implicated in misleading listeners. Non-standard consonants (singleton consonants and consonant clusters) were implicated in misleading listeners at 75 (85.2%) sites. The types of non-standard production of singleton consonants that misled listeners included substitution, where one was substituted for another (eg level → /levən/), addition, where a consonant was added (eg sitting → /sɪstɪŋ/) or a singleton consonant was added to a word final vowel (eg holiday → /hɒlideɪs/), and deletion, where a singleton consonant was missing (eg good → /gʊ/). The non-standard production of consonant clusters included reduction, where one or more of the consonants in the cluster was missing(eg entrance → /entrən/), substitution of one consonant in the cluster (eg after → /aptɜ/), reduction and substitution (eg child → /ʧaɪs/), substitution of all consonants in the cluster (eg three → /twi:/), deletion of the cluster (eg kind → /kaɪ/) and addition of a consonant (eg description → /dætskrɪpsən/).

The relative extent to which non-standard consonants in the initial, medial and final word positions were implicated at sites of RI is presented in Figure 1. As shown in Figure 1, non-standard consonants in the word final position appeared to cause the listeners the most difficulty, as they were implicated the most (at 56.8% of sites), followed by non-standard consonants in the medial position (at 28.4% of sites) and non-standard consonants in the initial position (at 19.3% of sites).

A closer look at the non-standard word final consonants represented in Figure 1 revealed that the non-standard production of 17 different single consonants and consonant clusters was involved. The speaker's non-standard production of consonant clusters containing /n/ and single consonants /s/, /t/ and /l/ seemed to cause the listeners the most difficulty, as they were each implicated at multiple sites and, together, accounted for 60% of the implicated word final consonants. The non-standard production of these consonants was as follows:

- Consonant clusters containing /n/ were most commonly produced in a reduced form with just the /n/ pronounced (n/ns, n/nt, n/nd), but at one site the entire cluster was deleted.
- /s/ was most commonly added to the ends of words, but at one site the speaker produced a /t/ in place of the /s/.
- /t/ was either deleted or another consonant produced in its place (eg s/t, d/t, n/t).
- /n/ was produced in place of /l/.

Non-standard vowels, other than those associated with a non-standard syllable stress pattern, were implicated in misleading listeners at 30 (34.1%) sites of RI. The relative extent to which non-standard vowels in strong and weak syllables were implicated is presented in Figure 1. It can be seen from Figure 1 that non-standard vowels in strong syllables were implicated to a greater extent (at 29.5% of sites) than those in weak syllables (at 5.7% of sites).

Thus, non-standard segments, particularly consonants, seem to have contributed significantly to reducing this speaker's intelligibility. Non-standard consonants in final position were particularly important, and non-standard vowels in strong syllables seemed to cause more problems than those in weak syllables. It should be noted, however, that although this analysis looked at the word position of implicated non-standard consonants, these consonants were produced in connected speech and, therefore, because there was no separation between the words at times, the implicated non-standard word final consonants were part of a continuous flow of speech; a situation which is different to the production of single words. They therefore had the potential to influence the listener's identification of the following word.

THE MIX OF NON-STANDARD FEATURES THAT REDUCED INTELLIGIBILITY

The way non-standard segments and non-standard features that rendered the syllable stress pattern non-standard combined to mislead listeners was often quite complex. At just over half of the sites (n=45; 51.1%) a combination of multiple non-standard features was implicated in misleading listeners, and at one particular site, seven different non-standard features were implicated (which means, of course, that only a single non-standard feature was implicated at just under half of the sites (n=43; 48.9%), that is, with a similar frequency). The extent to which non-standard syllable stress patterns and non-standard segments were implicated in misleading listeners is presented in Figure 2. As shown in Figure 2, at most sites where a non-standard syllable stress pattern was implicated, so, too, were non-standard segments (at 71.4% or 20 of the 28 sites). For example, in the following site from utterance 47, all three listeners identified the word *officer* as *official*, indicating that they were misled by a combination of the non-standard strength of second syllable being produced (and an associated non-standard full vowel) and the production of /ʃ/ as /s/.

officer or
ɒfɪʃə
s S w

It is also clear from Figure 2 that at the majority of sites (n=60; 68.2%) the syllable stress pattern was in fact standard, and non-standard segments misled the listeners. At 25 of these 60 sites, multiple non-standard segments were implicated in misleading listeners. For example, in utterance 23 all three listeners heard the word *postgraduate* as the words *both graduate*, indicating that they were misled by the combination of the production of /p/ as /b/ and /s/ as /t/.

I mean the postgraduate degrees
bəʊtgrædjuət
S s ww

Comments made by Listener 1 gave some insight into why the listeners may have identified the words as *both graduate*, rather than *boat graduate*, an exact replication of the speaker's production. Her comments suggest that she used her knowledge to compensate for the speaker's accent when attempting to identify his intended words. She transcribed the words *boat graduate* in her first transcription, but changed it to *both graduate* in her second, commenting: *I've changed* ***boat*** *to* ***both*** *because I don't think he can get his tongue around the 'th' sound. I think he's talking about graduate degrees. He's done two degrees so he graduates with both degrees.*

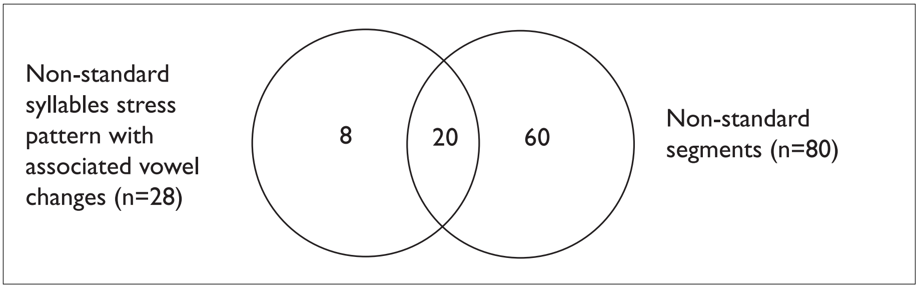
Non-standard syllables stress pattern with associated vowel changes (n=28)
8
20
60
Non-standard segments (n=80)

Figure 2: The number of sites of RI at which the speaker's production of a non-standard syllable stress pattern (with associated vowel changes) and/or non-standard segments were implicated in misleading listeners (n=88)

Conclusion and implications

For this speaker, intelligibility was reduced by a complex cocktail of different 'ingredients'. One ingredient was the speech processing strategies used by the listeners, who, by relying on these strategies, drew on familiar suprasegmental and segmental features of the speech signal to identify the speaker's intended words. The speaker's non-standard features added other ingredients into the mix, and these themselves comprised a complex mix of non-standard suprasegmental and segmental features that either resulted in an altered syllable stress pattern or non-standard segments. RI was the result of the interaction between the listeners' processing strategies and the non-standard features, both suprasegmental and segmental, in the speech signal. Where the syllable stress pattern was non-standard at an RI, it misled the listeners who appeared to rely heavily on it as a guide to understanding, but at most sites the syllable stress pattern was, in fact, standard. Thus, although the listeners did not seem to rely on the segments as a guide to the same extent as they relied on syllable stress pattern, the segments that were there in the speech input were non-standard at 90.9% of sites. Consequently, for this speaker, non-standard segments played a greater role in reducing intelligibility than did non-standard features in syllable stress pattern.

An understanding of how both listener and speaker contributed to this cocktail is important to our understanding of which non-standard features have an impact on intelligibility. If we were to consider only the listener ingredients we might conclude that the syllable stress pattern is of greater importance than the segments to intelligibility, because the listeners relied so heavily on it. Similarly, if we were to consider only the speaker ingredients we might conclude that segments are of greater importance than the syllable stress pattern, because they were implicated in misleading listeners at far more sites. Although for this particular Vietnamese speaker non-standard segments played a greater role in reducing intelligibility than did non-standard syllable stress patterns, if the listeners apply the same listening strategies when listening to a different speaker from a different language background[7] the situation might be different, the cocktail would be altered and non-standard syllable stress patterns may be more involved. Similarly, a listener from a different language background may draw on different features of this Vietnamese speaker's speech signal to identify his intended words and thus different non-standard features may be implicated in reducing intelligibility, as the listener ingredients of the cocktail would be altered.

As these findings relate to only one speaker and three listeners, they cannot necessarily be generalised to all Vietnamese L1-background speakers of English or all native English listeners. Previous research suggests that listeners from different backgrounds may draw on different aspects of the speech signal to identify a speaker's intended words (see note 4 below). However, before the strategies used by the listeners in this study can be attributed to the fact that they are native speakers of English, further research needs to investigate which features listeners from different language backgrounds would rely on in their attempts to identify this speaker's intended words, and whether or not they are the same as those relied upon by the listeners in this study.

The findings reported here support the findings of word recognition studies (Bond and Small 1992; Cutler and Butterfield 1992) that both the rhythmic properties of the speech signal and segments are important to English listeners. The listeners in this study relied on the syllable stress patterns and segments in the speech signals in their attempts to identify the speaker's intended words at sites of RI. The findings also add to the findings of previous studies (Suenobu, Kanzaki and Yamane 1992; Benrabah 1997; Bent, Bradlow and Smith in press) that, for English listeners, both non-standard suprasegmental and non-standard segmental features can influence intelligibility. Non-standard syllable stress patterns produced with the wrong number of syllables and/or containing syllables produced with non-standard strength were misleading to the listeners, as were non-standard vowels, particularly those in strong syllables, and non-standard consonants, particularly those in the word final position. Furthermore, many of the non-standard features that influenced intelligibility for this speaker are characteristic of the non-standard features found in English produced by speakers from a Vietnamese L1 background (Nguyen and Ingram 2004; 2005).

The finding that reduction in intelligibility may be the result of a complex mix of non-standard features of speech production is important when considering how to improve this speaker's intelligibility. Because of the wide range of non-standard features implicated and because multiple features were implicated at over half of the sites of RI, improving his intelligibility is not likely to be a simple task. Although the above findings indicate that word final consonants are an important issue for him, the elimination of this particular non-standard feature from his intelligibility 'cocktail' may not necessarily lead to immediate results. The process of improving his intelligibility may therefore be complex and need to incorporate a range of changes before results are evident. This finding highlights the need for further research to investigate the relative impact of the different non-standard features when more than one is implicated in misleading a listener.

This study, by allowing some access into the processes that occur when a listener transcribes an utterance, gives us further insight into the relationship between intelligibility and non-standard features of speech production. It adds more information to help us understand the way the different listener and speaker ingredients combine and interact with each other in the intelligibility cocktail. Hopefully further research will uncover new ingredients to add to the mix.

Appendix

EXAMPLES OF UTTERANCES

Because of space limitations, only the utterances containing those sites of RI used as examples in the article are included here. In each utterance the position of any pauses separating the pause groups have been marked with /.

Utterance 3	At that level / student have to / study for / five years.
Utterance 23	I mean the postgraduate degrees.
Utterance 47	They got a good / relationship with a / officer or / some well known teacher.

NOTES

1 Non-standard features of speech production are those that are different from what would be expected if the same section of speech was produced by a native speaker of (Australian) English.

2 A listener's ability to identify a speaker's intended words can depend on linguistic knowledge from a number of different sources. Grosjean and Gee (1987) described the process of word identification in connected speech as a 'feed-forward, feed-back system' which involves interaction between phonological, grammatical, situational and other types of information resulting in 'constant adjustments being made to early and/or partial analyses and constant prediction being made on what is to come' (Grosjean and Gee 1987: 148). Similarly, Moore (2003) supported the notion that the initial analysis of the phonological information is checked and readjusted using knowledge of the ways speech sounds follow each other, and adjusted and corrected further using knowledge of syntax and semantics, and even situational cues. This article, however, focuses on how the phonological information in the speech signal influences the listener's ability to identify a speaker's intended words.

3 Previous studies have investigated the influence of non-standard phonological features in L2 speech on listeners' ability to understand the content of what a speaker says (for example, Hahn 2004), and on pronunciation ratings using a scale that combined intelligibility and accent (Anderson-Hsieh, Johnson and Koehler 1992), but as the focus of this article is on listeners' ability to identify the words a speaker intended to say, only those studies that measure intelligibility in this way are included in this review. The focus of the research reported in this article is the intelligibility of connected speech in L2 speakers, and thus only the findings of previous research that focused on connected speech are included in this review. Words produced in connected speech can be somewhat different to the way they are produced as single words (see Shockey 2003). Consequently, the studies that investigated the influence of non-standard features on the intelligibility of single words (for example, Field 2005) have not been included. I have also included studies that investigated the intelligibility of sentences read aloud since some studies have found these to be comparable (see Munro and Derwing 1994).

4 Different features of speech may be important to non-native listeners listening to L2 speech (see Jenkins 2000; Jenkins 2002). Cutler, Dahan and van Donselaar (1997) and Cutler (2001) argued that the way listeners segment a stream of continuous speech into individual words is language specific and related to the listener rather than the speech signal. Cutler (2001: 11) asserted that when listening to their native language, listeners rely on 'strategies which are specifically tailored to the native phonology'. Listeners from different language backgrounds may therefore draw on different aspects of the speech signal to identify a speaker's intended words.

5 Because a listener's ability to understand an L2 speaker may vary according to his or her prior knowledge of the speaker's topic, prior experience in listening to non-native accents in general, as well as prior experience with the speaker's particular non-native accent (Gass and Varonis 1984), it was important that the three listeners were similar in this respect.

6 Those vowels associated with a non-standard syllable stress pattern were not counted here.

7 Findings from the larger study from which these results are taken indicate that the listeners applied the same listening strategies when attempting to identify the other L2 speakers' intended words (one from a Korean and one from a Mandarin L1 background).

REFERENCES

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, 44(5), 395– 408.

Benrabah, M. (1997). Word-stress – A source of unintelligibility in English. *IRAL*, 35(3), 157–165.

Bent, T., Bradlow, A. R., & Smith, B. (in press). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In M. J. Munro & O. S. Bohn (Eds.), *Second language learning: The role of experience in speech perception and production*. Amsterdam: John Benjamins.

Bond, Z. S. (1999). *Slips of the ear. Errors in the perception of casual conversation*. San Diego: Academic Press.

Bond, Z. S. (2003). *Slips of the ear*. Retrieved December 26, 2005, from http://oak.cats.ohiou.edu/~bond/Slips.htm

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics*, 34(5), 470– 474.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5(1), 1– 23.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218–236.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford: Oxford University Press.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2), 141–201.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1–16.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399– 423.

Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. Canberra: Department of Education, Training and Youth Affairs (Australian National Training Authority Adult Literacy National Project).

Gass, S., and Varonis, M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34 (1), 65–89.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25, 135–155.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201–223.

Hansen, J. G. (2004). Developmental sequences in the acquisition of English L2 syllable codas. A preliminary study. *Studies in Second Language Acquisition*, 26, 85–124.

Hwa-Froelich, D., Hodson, B. W., & Edwards, H. T. (2002). Characteristics of Vietnamese phonology. *American Journal of Speech-Language Pathology*, 11, 264–273.

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83–103.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllable strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America*, 104(4), 2457–2466.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego: Academic Press.

Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.), *Pronunciation pedagogy and theory. New views, new dimensions* (pp. 64–91). Alexandria, VA: TESOL.

Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11(3), 253–266.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97.

Nguyen, T., & Ingram, J. (2004). A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. *Proceedings of the tenth Australian international conference on Speech Science & Technology*. (pp. 516 –521). Sydney: Macquarie University.

Nguyen, T., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39(2), 309–319.

Shockey, L. (2003). *Sound patterns of spoken English*. Malden, MA: Blackwell Publishing.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111(4), 1872–1891.

Suenobu, M., Kanzaki, K., & Yamane, S. (1992). An experimental study of intelligibility of Japanese English. *International Review of Applied Linguistics*, XXX(2), 146–153.

Thompson, L. C. (1987). *A Vietnamese reference grammar*. Honolulu: University of Hawaii Press.

Wright, R., Frisch, S., & Pisoni, D. B. (1996–1997). *Speech perception* (Research on spoken language processing. Progress report 21). Bloomington, IN: Indiana University.