Intelligibility and the Listener: The Role of Lexical Stress

JOHN FIELD
Birkbeck College, University of London
London, England

For some 30 years, intelligibility has been recognized as an appropriate goal for pronunciation instruction, yet remarkably little is known about the factors that make a language learner's speech intelligible. Studies have traced correlations between features of nonnative speech and native speakers' intelligibility judgements. They have tended to regard prosody as a global phenomenon and to view intelligibility as primarily a quality of the speaker. The present article focuses on a single prosodic element, *lexical stress,* and shifts the focus of study to the listener. It draws on findings in psycholinguistics that have rarely been applied to second language (L2) contexts. Groups of listeners were asked to transcribe recorded material in which the variables of lexical stress and vowel quality were manipulated. Recognizing the extent to which English is employed in international contexts, the study contrasted the effect of the variables on native listeners (NLs) with their effect on nonnative listeners (NNLs). NLs and NNLs were found to respond in remarkably similar ways to the problems posed by stress misallocation. For both groups, the extent to which intelligibility was compromised depended greatly on the direction in which stress was shifted and whether changes in vowel quality were involved.

Arguably the most pressing issue in L2 pronunciation research today is the quest to identify the factors that most contribute to speaker intelligibility. Intelligibility is the target that pronunciation instruction traditionally sets for itself, but English language teachers know little about how best to help learners achieve it. A number of revealing studies have attempted to trace correlations between various features of nonnative speech and native speakers' intelligibility judgements. Characteristically, they have considered the following:

- complexes of deviations from native speaker norms (prosody, phonemic segments, accentedness) rather than individual factors
- speaker characteristics as manifested in selected samples of L2 English
- assessments of native speaker judges

This article focuses on a single factor, lexical stress, with a view to determining what emphasis it should be given in pronunciation instruction. Regarding intelligibility as a two-way process, it emphasizes the perceptions of listeners rather than the productions of speakers. It takes account of the extent to which English is today employed in international contexts by contrasting the effect of misplaced stress on NLs with its effect on NNLs.

The discussion is presented against a background of well-attested findings from cognitive psychology and speech science that have featured relatively little in accounts of second language acquisition (SLA). Underpinning the study is a view that the psycholinguistics of first language (L1) speech processing provides an important key to an understanding of the factors contributing to intelligibility.

INTELLIGIBILITY

The Construct

The notion of intelligibility has become central to the teaching of pronunciation. In 1949, Abercrombie famously remarked that "language learners need no more than a comfortably intelligible pronunciation" (p. 120). The idea was slow to feed through to practice, but in the 1970s many English language teachers worldwide came to recognize that it was unrealistic, time-consuming, and potentially inhibitory to aim for a native-like accent, and that such a goal might not necessarily represent the learners' wishes. They abandoned traditional checklist approaches to pronunciation instruction and instead adopted intelligibility as their criterion.

The first challenge lies in deciding how to define *intelligibility*. Much discussion of the construct has suffered from a failure to arrive at a clear consensus. Smith and Nelson (1985) note that terms such as *intelligibility* and *comprehensibility* are often used interchangeably. They suggest restricting the first to the recognition of word forms and utterances and the second to the construction of meaning. A similar distinction has been adopted by two major researchers in the field, Munro and Derwing (1995; Derwing & Munro, 1997), who apply it procedurally in their research methodology. Intelligibility is measured by the ability of judges to transcribe the actual words of an utterance, comprehensibility by an overall rating of how easy it is to understand a given speaker.

This division between local formal recognition and global processing effort makes sense when examining overall communicative success and failure. But Munro and Derwing's transcription task clearly embraces within intelligibility factors such as contextual transparency or syntactic

and lexical knowledge. It would not appear to correspond to the type of intelligibility specified in pronunciation teaching contexts as a desirable teaching goal; the latter is an altogether narrower construct involving the impact of strictly phonological factors on understanding.

This article consequently restricts the term *intelligibility* to features of the speech signal. As used here, it refers to the extent to which the acoustic-phonetic content of the message is recognizable by a listener. On this analysis, intelligibility forms part of a wider construct of comprehensibility.

The distinction helps to position the present study within an area of specifically phonological enquiry. It also serves to separate perceptual evidence at phoneme, word, and tone-group levels from higher level evidence such as world knowledge, which originates outside the signal.

Contributory Factors

A major challenge for pronunciation specialists is to identify which features of natural speech contribute most significantly to intelligibility. This information is critical if they are to set priorities and to devise principled programs of instruction. Gimson (1978) initiated discussion on this issue by proposing a simplified phonological system (*rudimentary international pronunciation*) for nonnative speakers, which accepted modifications to certain problematic sounds (principally, voiced consonants and diphthongs) on the grounds that they would not greatly affect the speaker's ability to be understood. The sixth edition of Gimson's pronunciation guide (1994, pp. 283–287) makes more concrete suggestions as to the tolerances that might be acceptable when aiming for *minimum general intelligibility*. It proposes a vowel system of six short vowels, seven long vowels, and three diphthongs, but the proposals for consonants license only minor deviations from native-speaker norms.

A more radical approach to the issue of pronunciation course content is found in Jenner's (1989) proposal that there may be a *common core* of phonological features that, if taught systematically, would establish a framework for intelligible speech. Jenkins (2000) extends this view, suggesting that a new international form of English may evolve that retains those features most critical to intelligibility between nonnative speakers but suppresses others that are peripheral.

Segmental vs. Suprasegmental Features

It is by no means easy to determine which features of pronunciation should be prioritized on the grounds that they enhance a learner's

intelligibility. In particular, opinion has been divided as to the relative contribution made by segmental features (phonemes) and suprasegmental ones (word stress, rhythm, and intonation, often referred to collectively as *prosody*).

The present study focuses on suprasegmentals. There are two principal reasons for this decision. Firstly, research evidence suggests that suprasegmentals play a more important role than segmentals. Anderson-Hsieh, Johnson, and Koehler (1992) compared the relative contributions made to intelligibility by prosody, segmentals, and syllable structure. Within 11 different language groups, they found that the score for prosody was most significantly associated with the overall score for pronunciation. A similar finding was reported by Anderson-Hsieh and Koehler (1988), who concluded that "prosodic deviance may affect comprehension more adversely than does segmental deviance" (p. 562). In a related finding, Derwing, Munro, and Wiebe (1998) studied the effects of both segmental and suprasegmental instruction on learners' comprehensibility ratings and concluded that the latter had a greater effect on performance in communicative contexts.

A possible explanation for the low impact of segmental errors is found in psycholinguistic accounts of first language (L1) processing. Commentators such as Marslen-Wilson (1987, p. 95) have suggested that an L1 listener could only succeed in finding a match for a mispronounced word such as *shigarette* if the process operated on a principle of best fit rather than exact match. In the context of the intelligibility debate, this means that the occasional insertion of a nonstandard phoneme should not grossly disrupt communication.

The second consideration was a methodological one. Studies of L1 listening (e.g., Elman & McClelland, 1988; Ganong, 1980) have demonstrated that lexical knowledge plays an important part in how a listener processes a group of phonemes, especially where the signal is imprecise. This finding means that listening researchers have difficulty disentangling the effects of phoneme error from those of whole-word matching. Suppose that an NL has no problem understanding a nonnative speaker who produces the word *veshtables*. One might conclude that the substitution of /ʃ/ for /ʧ/ does not impair intelligibility, making /ʃ/ a low-priority item for the teacher. However, an equally valid interpretation would be that the listener had drawn on his or her knowledge of the existence of the word *vegetables* and applied a best fit strategy. Or, more likely, the truth might lie in an interaction between the two sources of evidence (McClelland & Elman, 1986). This perspective posed practical problems at a segmental level for the kind of transcription study projected for this article; it was felt that a feature at lexical level or above would prove a more manageable target.

Lexical Stress

A number of studies (e.g., Anderson-Hsieh et al., 1992; Anderson-Hsieh & Koehler, 1988; Derwing et al., 1998) have sought to assess the impact of suprasegmental features on intelligibility by contrasting it with the impact of other factors such as phoneme accuracy or accentedness. They have tended to treat prosody as a unitary construct. However, it seems probable that the various constituents of prosody (lexical stress, intonation, the relative duration of weak and strong syllables) contribute to intelligibility in different ways. The present study focuses principally on lexical stress, though it also considers the variations in vowel quality and syllable duration that are associated with it. There are a number of reasons for believing this particular feature to be important.

Firstly, lexical stress plays a central role in determining the profiles of words and phrases in current theories of metrical phonology (Hogg & McCully, 1987). Secondly, psycholinguistic studies of *slips of the ear* (Bond, 1999) have shown that native English listeners place greater reliance on the stressed syllables of words than on the briefer unstressed syllables. When native speakers are asked to *shadow* (repeat back) speech containing pronunciation errors, they are three times more likely to detect and reproduce an example of misplaced stress than one of a mispronounced phoneme (Bond & Small, 1983).

Thirdly, some accounts of speech processing (e.g., Grosjean & Gee, 1987) raise the possibility that the stressed syllable of a word provides the listener with a code that links directly to the representation of the word in the mind. On this hypothesis, the syllable /næ/ guides the search for the word *international* as does the syllable /tɒg/ for the word *photography*.[1] Applying this notion to an L2 context, differences might be observable between the importance accorded to stressed syllables by a native speaker with long-established procedures for locating words and that accorded by a nonnative speaker who has not yet fully established a set of appropriate codes. Some speakers may not ever acquire such codes: Peperkamp and Dupoux (1992) suggest that speakers of fixed-stress languages do not even store lexical stress as part of their phonological representation of words.

The most compelling reason for investigating lexical stress lies in an L1 study that produced striking evidence that certain types of stress misplacement appear to seriously impair intelligibility. Cutler and Clifton (1984) switched the stressed syllable in disyllabic words such as *canTEEN*

[1] A stumbling block for this theory is the relative nature of stress as explored in metrical phonology. How would a listener respond to the secondarily stressed [pek] in *expectation* given that it is heard before the syllable [teɪ], which potentially forms the access code? Might it trigger a search among words such as *impeccable* or *respectable*?

and *TURbine* and, using reaction time measurements, studied the extent to which the substitution impaired the ability of NLs to identify the words. They reported no effect on intelligibility in the case of a leftward shift of stress (as in words like *CANteen*.) However, intelligibility was somewhat impaired when the shift was rightward (as with *turBINE*); and, most importantly, intelligibility was seriously compromised when the shift of stress also entailed a change of vowel quality, as it might in a switches such as *laGOON* → *LAgoon* ([lə'guːn] → ['lægun]) or *WAllet* → *waLLET* (['wɒlɪt] → ([wɒ'let]).

Cutler and Clifton studied word forms in only two conditions: one with correct stress placement and one with both stress and quality shifted. However, an interim condition is conceivable where these words suffer a shift of stress without an associated change of quality. The English phonological system restricts schwa to weak unstressed syllables, but that does not eliminate the possibility of a variant such as [wɒ'təː] for *WAter* in the speech of a nonnative speaker. It is even easier to envisage instances where stress might be added to syllables featuring the other weak quality vowels /ɪ/ and /ʊ/, forming variant forms such as ['ɪnʤɔɪ] (*ENjoy*) or ['tʊmɒrəʊ] (*TOmorrow*). So the Cutler and Clifton findings can be extended by examining the effects on intelligibility when a weak quality syllable is accorded stress with no change of quality. This study also extends the enquiry to NNLs to discover if they suffer the same loss of intelligibility as NLs.

METHODOLOGY

The most widely adopted approach to researching intelligibility (Derwing & Munro, 1997; Munro & Derwing, 1995) asks NL judges to evaluate samples of nonnative speech for prosody, accentedness, and other features, and then to rate them objectively for intelligibility—often on the basis of how accurately they have been transcribed. The present study adopts a very different approach. It employs a single set of exemplars in which two linked variables (lexical stress and vowel quality) have been manipulated.[2] The material is played not only to NL judges but also (with international intelligibility in mind) to a group of NNL judges.

The point should be made that this is a controlled experimental study whose purpose is to focus on a single phonological feature. In this respect it differs from earlier work, which has considered the relative contributions of a range of factors. The research question strictly concerns whether lexical stress does or does not contribute to intelligibil-

[2] The approach of Hahn's (2004) study of sentence stress and intelligibility is not dissimilar.

ity, and, if it does, in what way. There is, of course, no implication that lexical stress is the only factor that contributes to the construct.

Materials

Two groups of 12 disyllabic English words were used (Table 1). The first group followed a SW (strong-weak) pattern in standard British English (e.g., *SECond*). The second followed a WS (weak-strong) pattern (e.g., *beGIN*). The words in both groups were chosen because they were of a frequency higher than 100 per million in the British National Corpus (Leech, Wrayson, & Wilson, 2001)[3] and were likely to occur at an early stage in any course of L2 instruction. Of the WS items, a number had initial syllables that resembled prefixes (*return, prevent*) but none was a true prefix in the form of a separable morpheme (as in *re+visit* or *pre+historic*).

Within each group, several words were identified where a shift of stress might be expected to lead to a change of vowel quality. The criteria for determining the probability of this happening were

- the derivational morphology of the item in question (*woMEN*), or
- the existence of analogous words (*seCOND* as in *secondment*), or
- a change of vowel quality in a contrastive stress situation (*I said "CONtain" not "DEtain"*).

TABLE 1
Disyllabic Words Employed in the Study

SW Words	WS Words
husband	enjoy
second	forget
different	contain
person	provide
Friday	today
women	
follow	begin
coffee	decide
water	discuss
listen	inform
money	perhaps
notice	prevent
	return

[3] Figures were taken from the whole corpus rather than the spoken section since the NNL participants had had a comparatively short exposure to everyday speech. The exception to the 100 limit was *inform* (58), preferred to the more frequent *include* (353) because it preserves the [n] allophone in the first syllable.

In all, six words were selected from the SW group and five from the WS group. They form the first subset in Table 1. These words were to be recorded in three conditions:

1. standard form; example: ['sɛkənd].
2. shift of stress with no change of weak vowel quality (S); example: [sɛ'kəːnd].
3. shift of stress with change to full vowel quality (S + Q); example: [sɛ'kɒnd].

The remaining 13 words fell into two types: those without a weak quality syllable so that a change of quality was not possible (e.g., *FOllow → foLLOW*) and those with a weak quality syllable that would remain the same even if accorded stress (example: *LISten → listEN*). These words were only to be recorded in the first two conditions (standard form and stress shifted). Table 2 makes clear how the material was distributed.

This gave a set of 59 items, to which one more (*about* in standard form) was added to make the number 60. The standard forms of the target words were included alongside the stress-shifted variants for two reasons. Firstly, they acted in effect as foils, ensuring that the subject treated the test as a transcription exercise involving actual words rather than nonwords. Secondly, they provided a baseline for the normal intelligibility of these words in isolation, against which could be measured any decline in intelligibility when the words occurred in stress-shifted form. The baseline was particularly important in the case of the NNLs. It was reasonable to assume that they knew the target words, but that did not necessarily entail that they would recognize them in their spoken form.

A male native speaker of British English with no knowledge of phonology or background in language teaching was asked to record the items, following simple orthographic cues that indicated where stress or vowel quality shifting was involved. Each item was spoken five times. The recording was made digitally on a computer using a studio-standard microphone.

The five different versions of each item were then analyzed using a Soundblaster Audigy 2 ZS Platinum Pro editing program (Creative

TABLE 2
Distribution of Items Across Conditions of Lexical Stress

Condition	SW words	WS words
Standard	12	12
Stress-shifted	12	12
Stress + vowel quality shifted	6	5

Technology, 2003) operating at 16 bits and 22kHz. The aim was to choose a set of exemplars that were as consistent as possible. Three important acoustic factors influence the perception of stress in English: *intensity, duration,* and *pitch movement* (Fry, 1958). Special attention was given to duration because it is a major indicator of weak versus full vowel quality (Crystal & House, 1990), and it is the variable easiest to monitor. The duration of each stressed syllable was checked so as to give preference to those exemplars where it was closest to a ratio of 1.5 times the duration of the unstressed syllable (the mean for English as identified by Delattre, 1965). Once a provisional set of 60 exemplars had been identified, each was further checked for the relative level of intensity of its stressed syllable. The final set of items was submitted to two judges with phonetic training, who confirmed that the relevant syllable in each bore unambiguous stress marking and that there had been no shifts in vowel quality in the S (stress shift only) items.

The acoustic-phonetic content of the recorded materials might arguably have been controlled more rigorously by using synthesized speech or splicing together pieces of connected speech. However, the relationship between stressed and unstressed syllables is a complex one, the product of extremely fine timing decisions by the speaker that closely reflect his or her current speech rate (Janse, Nooteboom, & Quené, 2003). It was therefore considered preferable to record natural utterances by a naive speaker and to select good exemplars from among them.

Presentation of Stimuli

The material was presented as single words rather than embedded in context-neutral sentences such as "The next word you will hear is . . . ". This approach leaves the study vulnerable to the suggestion that the items might have been harder to recognize if they had occurred in running speech. However, whole-sentence contexts can compromise the validity of transcription tasks by allowing participants time to reflect on and revise what they have written. Experience also suggests that neutral carrier sentences often lead speakers in any case to produce word forms resembling citation ones—especially when, as here, they have to modify a standard pronunciation.

The 60 items were randomized into 3 sets of 20 by drawing lots, but in such a way that each set contained only one version of a given word. In addition, the three conditions were distributed as evenly as possible, so that each set contained 8 S items out of 24 and at least 3 S + Q items out of 11. There were never more than 2 consecutive items representing the same condition. Two additional words, *thousand* and *expect,* were inserted

at the beginning of each set to enable participants to normalize to the speaker's voice and to accustom themselves to the transcription task. The 3 sets of items were then recorded on to CD-ROM and finally on to cassette.

Participants

The materials were played to two groups of participants: one NL and one NNL. The NL participants were pupils at a British secondary school whose mean age was around 15. They were tested in three classes of similar ability levels (N = 28, N = 28, N = 26). Each class was asked to listen to and transcribe one set of the recorded items. The tests were carried out by the normal class teachers. None of the participants reported any hearing difficulties.

The NNL participants were students of English at two leading British private language schools: International House London and Eurocentre Cambridge. They were in 12 separate classes; four (N = 30) were tested on Set A of the recorded items, four (N = 22) were tested on Set B and four (N = 25) were tested on Set C. The participants were controlled for level; all had been tested on entry by their school and classified as intermediate. They represented a range of first languages. The principal L1 groups were Korean (N = 16), Japanese (N = 15), Mandarin Chinese (N = 10), Spanish (N = 9), Portuguese (N = 6), and Italian (N = 6). Results for each of these groups are examined independently. Other languages were German (N = 4), French (N = 2), Arabic (N = 3), Farsi (N = 1), Russian (N = 1), Polish (N = 1), Czech (N = 1), Georgian (N = 1), and Bulgarian (N = 1). The script of one Arabic speaker in Set C was rejected because the spelling was hard to interpret. This reduced Set C to 24 and left a total NNL population of 76.

Each NNL was asked to specify how many years he or she had studied English. Information was sought on the date of arrival in the United Kingdom; most participants had arrived no more than 3 weeks before the test and none had arrived more than 8 weeks before. No participant had previously spent an extended period in an English-speaking country. All participants were informed of the nature of the research and appropriate consents were obtained.

Procedure

Participants were in their normal class groups and in rooms with good acoustics (in the case of the NNLs, the rooms were specially designed for language teaching). One version of the materials was played to each class. With the NLs it was played by the class teacher under instructions

provided by the researcher; with the NNLs, the test was sometimes conducted by the class teacher and sometimes by the researcher. Good quality, steady-state cassette players were used.

The participants were given an answer sheet on which to provide personal details and to write their transcriptions. The instructions on the answer sheet were worded as follows:

> You will hear a voice saying a number followed by a word. Try to write down what you hear. Altogether, you will hear 22 English words. You will know some of the words, but some may be new to you. Try to guess the spelling of the word even when you do not recognize it.

The test administrator gave similar oral instructions. The intention was to leave open the possibility that the recorded items might represent known words or words that fell outside the participants' current vocabulary.

The scripts were analyzed, making due allowance for possible orthographic uncertainties. Given the proficiency level of the NNL participants, it proved possible throughout to determine where the respondent was aiming for a known word but had slightly misspelled it and where the target item was regarded as a new word. The one exception was the Arabic speaker in Set C whose script was rejected.

RESULTS

NL Responses

NL erroneous responses were calculated by participants. Z-tests revealed a significant difference in all three experimental groups between the level of error when transcribing a word in its standard form and the level when transcribing it with shifted stress (Set 1: $z(1) = 6.04$, $p < 0.001$; Set 2: $z(1) = 6.38$, $p < 0.001$; Set 3: $z(1) = 2.94$, $p < 0.001$). However, only one group showed a significant difference between the level of error for standard form items and the level for those where both stress and quality had been modified (Set 1: $z(1) = 0.68$, $p = 0.50$, n.s.; Set 2: $z(1) = 1.95$, $p < 0.05$; Set 3: $z(1) = 0.37$, $p = 0.71$, n.s.). Low standard deviations in the two shifted conditions indicated a low level of variation between participants.

The NL responses were then classified by items according to whether the target word had been recognized. Table 3 shows the results for, respectively, the 20 items in their standard form, the same items with stress shifted and a subset of the same items with both stress and vowel quality adjusted.

A chi-square test for independence indicated that the differences between responses in the three conditions were highly significant: $\chi^2 (2) =$

TABLE 3

NL Correct Responses Across Three Conditions of Lexical Stress

Condition	Correct	Total	%	Mean per item	SD
Standard (N = 24)	627	656	95.58	26.12	2.13
Stress shifted (N = 24)	503	656	76.68	20.96	6.66
Stress + vowel quality shifted (N = 11)	267	302	88.41	24.27	2.53

101.80, $p < 0.001$. Level of recognition of the target items in their normal form was then contrasted with recognition in the two variant conditions. Significant differences were confirmed by z-tests. For shifted stress, $z(1) = 3.31$, $p < 0.001$; for shifted stress and quality, $z(1) = 2.50$, $p < 0.01$.

Overall, stress shifting, with or without an accompanying change of vowel quality, was found to impair intelligibility. However, the impact of changing both stress and quality emerges in this data as less of a threat to intelligibility than the effect of shifting stress alone. This finding conflicts with that of Cutler and Clifton (1984).

An attempt was made to quantify the overall loss of intelligibility that resulted from the changes in lexical stress. The figures were derived by subtracting total percentage recognition in the nonstandard conditions from total percentage recognition of the same items when presented in standard form. The difference was then quoted as a proportion of the standard form figure. The resulting figures suggested overall decrements in intelligibility of 19.78% for S variants and 7.50% for S + Q variants.

The items were then grouped according to whether the nonstandard forms involved a rightward or a leftward shift of stress. Again, intelligibility of the standard form was compared with intelligibility of the variant forms. The results for the S condition are shown in Table 4 and for the S + Q condition in Table 5. The z-test figures indicate the statistical significance of the differences between the standard-form condition and the shifted ones.

A rather more complex picture now emerges. The degree to which intelligibility is impaired appears very much to reflect the direction of the stress shift. When stress is shifted leftward, the impact is considerably less than when it is shifted right. Strikingly, when it is shifted leftward with an accompanying change of vowel quality, it does not lead to any statistically significant reduction in intelligibility.

This finding was checked by comparing the results for the S conditions with the results for the S + Q conditions. With right-shifted items,

TABLE 4

NL Responses by Direction of Shift: Stress-Shifted Condition

Condition	Correct	Total	Mean per item	SD	Significance
Standard (N = 12)	312	328	26.00	1.86	
Left shifted (N = 12)	287	328	23.92	2.19	$z(1) = 2.19$, $p < 0.05$
Standard (N = 12)	315	328	26.25	1.86	
Right shifted (N = 12)	233	328	19.42	6.81	$z(1) = 3.35$, $p < 0.001$

the difference was significant ($z(1) = 3.13$, $p < 0.001$); with left-shifted items, however, it did not reach significance ($z(1) = 1.29$, $p = 0.20$, n.s.). Even within the S condition, there was a significant difference between the effects of right and left movement. Results for right-shifted items (N = 12) showed that intelligibility was more extensively impaired than with left-shifted items ($z(1) = 2.18$, $p < 0.05$).

NNL Responses

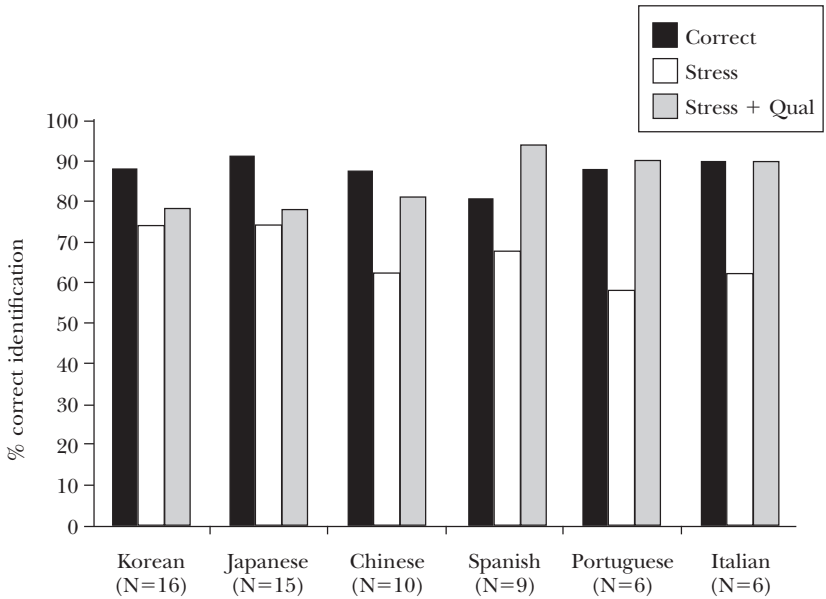
The NNL participants covered a range of L1s. The results were therefore first analyzed by subject, then combined into language groups. The six groups showed a considerable degree of consistency (see Figure 1). For all six groups, the effect on intelligibility of shifting stress alone was greater than the effect of shifting stress and vowel quality (though the differential between the two variant conditions was smaller with the Korean and Japanese groups). A curious finding was that for the Spanish

TABLE 5

NL Responses by Direction of Shift: Stress + Vowel Quality Condition

Condition	Correct	Total	Mean	SD	Significance
Standard (N = 5)	132	136	26.4	1.67	
Left-shifted (N = 5)	131	138	26.2	1.79	$z(1) = 0.18$, $p = 0.86$, n.s.
Standard (N = 6)	160	164	26.67	2.42	
Right-shifted (N = 6)	136	164	22.67	1.86	$z(1) = 3.21$ $p < 0.001$

FIGURE 1
Effects of Shifting Stress and Vowel Quality: Six Language Groups



group, the change of stress and vowel quality appeared to enhance intelligibility.[4]

A chi-square test of differences compared the correct responses of the six language groups across the three conditions. The result, χ^2 (10) = 4.38, p = 0.99, n.s., supported the null hypothesis of no significant difference between the groups.

This particular finding was checked using an analysis of variance, on the assumption of a normal population distribution.[5] Nine members were randomly chosen from each of the four largest language groups. Their results were analyzed across the three conditions with scores for the S + Q condition standardized by doubling them. A two-factor repeated measures ANOVA indicated a main effect of stress placement (standard, S, or S + Q): F (2, 107) = 10.95, $p < 0.001$. The participant's

[4] Too much should not be made of this result, given the small number of participants (N = 9). However, the effect chiefly occurred with a leftward shift of stress and full vowel quality. These conditions might have assisted speakers of Spanish, a language where penultimate syllable stress predominates and where prefixes are not relegated to weak syllables of short duration as they are in English.

[5] Thus avoiding the dangers of a familywise error rate if the chi-square test were to be repeated on subsections of the same data.

L1 was not a significant factor: $F(3, 107) = 0.47$, $p = 0.70$, n.s. However, there was a borderline interaction between the L1 and the effects of stress placement: $F(6, 107) = 2.18$, $p = 0.052$.

It would thus appear that (at least in respect to these materials) the major factor determining the results was the nature of the acoustic-phonetic signal rather than the participant's L1. This result is perhaps unsurprising. At their level of proficiency, the participants might be expected to have developed a degree of sensitivity to the acoustic-phonetic cues marking English stress. Furthermore, stress is a widespread phenomenon in the languages of the world. Some of the features that mark it (intensity and duration) contribute importantly to the processing of nonspeech signals such as music and would appear to be universals of the human auditory system.

Henceforth, the NNL participants were treated as a single group. Table 6 summarizes the responses by items of all NNL participants ($N = 76$) across the three conditions. Like the NL responses in Table 3, they show a marked decrement in intelligibility in the S condition and a less marked one in the S + Q condition. A chi-square test for independence showed a highly significant difference between the three conditions: $\chi^2(5) = 72.69$, $p < 0.001$. Using z-tests, participants' recognition of the target items in their normal form was then contrasted with recognition in each of the two stress-shifted conditions (S and S + Q). The differences were found to be significant. For shifted stress, $z(1) = 3.01$, $p < 0.001$; for shifted stress and quality, $z(1) = 11.38$, $p < 0.001$.

As with NLs, an attempt was made to quantify the loss of intelligibility that resulted from the changes in lexical stress. Using the same method, the overall decrement in intelligibility in the S condition was calculated at 21.28%, while the decrement in the S + Q condition was 7.10%. These figures are remarkably similar to those reported for the NNL group (19.78% and 7.50%, respectively), even though the NNLs started from a lower baseline in that their recognition of items in their standard form was less accurate than that of the NLs.

TABLE 6

NNL Correct Responses Across Three Conditions of Lexical Stress

Condition	Correct	Total	%	Mean	SD
Standard (N = 24)	530	608	87.17%	22.08	4.49
Stress shifted (N = 24)	422	608	69.41%	17.58	5.78
Stress + vowel quality shifted (N = 11)	229	280	81.79%	20.82	5.25

The results were then subdivided to examine the effects on intelligibility of shifting the stress to the left as against the effects when it was shifted to the right. Using recognition of the standard forms as a benchmark, Tables 7 and 8 show intelligibility under the two main conditions. The z-test figures in the final column indicate that, as with the NL results, the difference between the recognition of standard forms and the recognition of right-shifted forms reaches statistical significance, but that the difference between the recognition of standard and of left-shifted forms does not.

Thus striking parallels emerged between the levels of identification achieved by the NLs and those achieved (from a lower base) by the NNLs. Figure 2 shows intelligibility of the items in standard form compared with their intelligibility when lexical stress was manipulated. It shows that, for both groups of participants, intelligibility was reduced considerably more by shifting stress rightward without any change of quality than by shifting it leftward. The loss of intelligibility was lower when the stress shift was accompanied by a change of quality, but in this instance also, a leftward shift had less impact than a rightward one.

DISCUSSION

Findings

The study affords a number of possible insights into how lexical stress placement contributes to intelligibility. Firstly, it demonstrates a significant decrement in intelligibility when stress is shifted to an unstressed syllable without an accompanying change of quality. This finding held as true in the case of items such as *follow,* where the stress was shifted to a full-quality syllable (a decrement for NL participants of 48.15%), as in

TABLE 7
NNL Responses by Direction of Shift: Stress-Shifted Condition

Condition	Correct	Total	Mean per item	SD	Significance
Standard (N = 12)	275	304	22.92	3.58	
Left shifted (N = 12)	240	304	20.00	5.77	$z(1) = 1.49$, $p = 0.14$, n.s.
Standard (N = 12)	255	304	21.25	5.28	
Right shifted (N = 12)	182	304	15.17	4.88	$z(1) = 2.93$, $p < 0.01$

TABLE 8

NNL Responses by Direction of Shift: Stress + Vowel Quality Condition

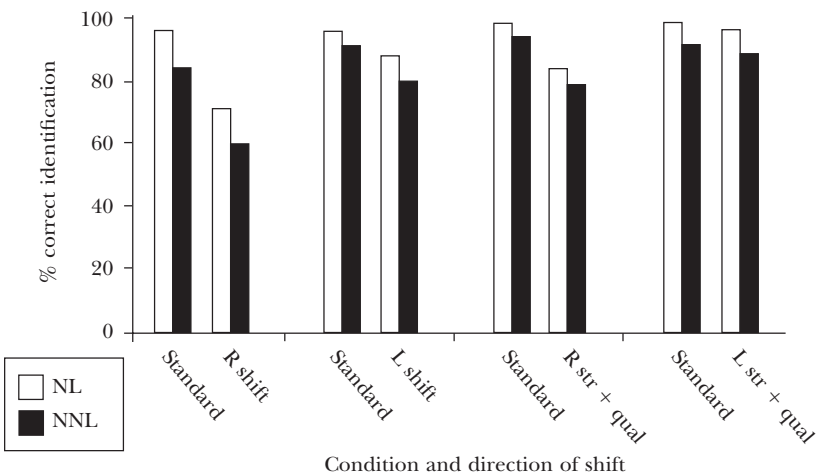
Condition	Correct	Total	Mean per item	SD	Significance
Standard (N = 5)	110	122	25.33	3.45	
Left-shifted (N = 5)	111	128	25.33	3.50	$z(1) = 0.06$, $p = 0.95$, n.s
Standard (N = 6)	141	152	23.5	3.45	
Right-shifted (N = 6)	118	152	19.67	3.50	$z(1) = 1.91$, $p < 0.05$

the case of items such as *listen,* where it was shifted to a weak-quality syllable (a decrement of 25.93%). However, where the stress shift was accompanied by a change of vowel quality (from weak to full), the loss of intelligibility was considerably less marked.

The finding appears to run counter to that of Cutler and Clifton (1984), although their study differed in two important ways. It did not include cases where stress was shifted to weak quality syllables with no change of quality (['wɪmɪn] → [wɪ'mɪːn]). In addition, Cutler and Clifton used reaction time data to measure difficulty in recognizing the non-standard forms, whereas this study concerned itself with the proportion of cases in which intelligibility was entirely lost.

FIGURE 2

Intelligibility by Items: NL vs. NNL



One possible explanation of the finding is that the loss of intelligibility in the case of NLs was attributable to the novelty of hearing a weak-quality vowel in a stressed syllable. But fewer than half of the target words had schwa in their weak syllable; others featured weak quality /ɪ/ or /ʊ/ vowels, which certainly occur in lexically stressed contexts (*winter, football*).[6] A more likely explanation lies in the finding (Bond & Small, 1983) that NLs (and by extension experienced NNLs) learn to place faith in the *island of reliability* afforded by the lexically stressed syllable of a word. It has even been suggested, as noted earlier, that such syllables provide access cues to words in the listener's mental word store. It is therefore to be expected that processing will be disrupted if

1. the stressed syllable contains a vowel of low informativeness (Gimson, 1994, p. 136), such as the high frequency schwa or /ɪ/.
2. the stressed syllable does not provide an access cue to the appropriate set of lexical candidates: Stressed /ɪ/ in *woMEN* would link to *diMINish, MINute,* and possibly even stored chunks such as *coME IN.*

So why the difference when the shift of stress is accompanied by a change of quality? One explanation is that a substituted full vowel is more informative than schwa or /ɪ/ and is also likely to be longer and thus more perceptually reliable. Another is that the full-quality syllable often bears a close relationship to the orthographic form of the word and may thus (at least for NNLs) provide a clearer and more reliable cue than the short, weak-quality one it replaces. It is interesting that substitution also assisted recognition by NLs, suggesting that this kind of orthographic link contributes to their processing, too.

The second major finding was that intelligibility is much more frequently impaired when lexical stress is shifted to the right than when it is shifted to the left. This finding confirms a similar one by Cutler and Clifton (1984). One reason may be that English specifically licenses a leftward shift of lexical stress in certain circumstances, especially for purposes of contrast:

> *I said INform them, not REform them.*
>
> *Instead of CONtaining the rioters, they DEtained them.*

It thus seems likely that our phonological representation of items such as *inform* or *contain* allows some degree of latitude to accommodate these stress-shifted forms.

Similarly, some of the WS items featured in the study had initial weak syllables that correspond exactly to frequent function words (*for-* in *forget, con-* in *contain, be-* in *begin, to-* in *today*). The words in question have two

[6] Though admittedly not in an open syllable, as in *money.*

forms, a weak unstressed one and a full one capable of bearing stress; listeners have to learn to accept either. A process of analogy may extend this tolerance to prefixes and prefix-like syllables that are auditorily identical to functors.

Within the two groups, the responses were very consistent. One of the more striking findings was the extent to which results in the NNL group were sustained across L1 divisions. This finding suggests that, at this level of proficiency and regardless of L1, learners have acquired the ability to recognize and exploit the cues provided by lexical stress.

There was also a consistent between-groups pattern of behavior, with NLs and NNLs responding to nonstandard lexical stress in remarkably similar ways. Although NNLs began from a lower base in terms of recognizing items in standard form, the two groups manifested a similar loss of intelligibility across the two variant conditions. This finding ran counter to an early hypothesis that NNLs might have become habituated to the uncertain stress placement of their peers and thus might show themselves significantly more tolerant than NLs of deviations from the stress norm.

CONCLUSION

English lexical stress does not normally serve to distinguish between lexemes: Cases such as *FOREbear/foreBEAR* cited by Cutler (1986) are relatively rare.[7] The issue at stake in this study is therefore purely and simply whether incorrect placement of lexical stress by a nonnative speaker renders the form of words unintelligible to an interlocutor.

Attempts to quantify the effects of misplaced stress suggested an overall decrement of 19.78% for NLs and 21.28% for NNLs. Cutler and Carter (1987) have calculated that polysyllabic items of the type studied here constitute 40.59% of all words in English conversation. This finding might suggest that the threat to intelligibility posed by incorrect placement of lexical stress is, relatively speaking, quite small: affecting only around 8% of content words if every word were misstressed. However, it should be borne in mind that the items tested here were in citation form; they would presumably be much harder to identify if occurring in running speech, where word pronunciation is more subject to variation.

Furthermore, the consequences of misinterpreting even a small number of content words can be extremely damaging to global understanding. Some EFL/ESL practitioners hold to the view that perceptual

[7] Stress certainly serves to identify word class in a limited set of about 300 noun-verb pairs (*an EXport, to exPORT*), but this characteristic of the English lexicon is unlikely to lead to any breakdown of understanding caused by a misattribution of meaning.

errors are relatively trivial because listeners can compensate for them by drawing on information provided by context in the form of the listener's understanding of what has been said so far. This argument is circular, however, because *context* in this sense depends entirely on how much of the previous input the listener has been able to decode accurately. To give an example, if the misstressed item *foLLOWED* occurs toward the beginning of an utterance, it might well lead the listener to construct a mistaken meaning representation around the notions of *load* or *flowed;* this representation would then shape the listener's expectations as to what was likely to follow. Evidence is emerging (Field, 2004) that NNLs place great reliance on interpretations at word level, even in the face of contradictory evidence.

One might conclude, then, that lexical stress should be an area of concern for pronunciation teachers, though perhaps not a top priority. However, two limitations of the present study indicate that stress allocation is potentially more important than has been suggested.

Firstly, the study featured a postperceptual recognition task: The transcription gave participants time after hearing the recording to form conclusions about what they had heard. A further study is needed that investigates the extent to which misplacing lexical stress increases the listener's processing load at the moment of hearing. Let us accept, for the sake of argument, the notion that the stressed syllable of a word provides an access code or, at the very least, a reliable signpost to its identity. Shifts in stress will then clearly create *garden path* situations: For example, stressing the second syllable in *foLLOW* will lead the listener toward a cohort (Marslen-Wilson, 1987) that includes *low, local,* and possibly *below,* and away from the target word. The consequent increase in processing demands might well limit the listener's ability to perform under the pressures of a conversational context.[8]

Secondly, the present study focused on single words in isolation. But one of the major functions of lexical stress is to assist listeners in dividing up whole stretches of connected speech by providing cues as to where words begin and end. In many languages (Hyman, 1977), this function is achieved by a fixed stress on the first or last syllable of a word. In English, NLs appear to employ a *metrical segmentation strategy* (Cutler & Norris, 1988) which exploits the fact that around 90% of content words in connected English speech are either monosyllabic or bear lexical stress on their first syllable (Cutler & Carter, 1987). By working on the assumption that each stressed syllable[9] marks the onset of a new word,

[8] See Hahn (2004) for garden path evidence on sentence stress.

[9] The term *stressed* as used here is somewhat of a simplification. Cutler and Norris refer to *strong* syllables, by which they mean syllables not bearing weak vowel quality. However, other commentators (e.g., Grosjean & Gee, 1987) use stress as their criterion.

NLs are able to divide up the speech stream with a fair degree of accuracy.[10] Thus, if lexical stress is wrongly distributed, it might have serious consequences for the ability of the listener, whether native or nonnative, to locate words within a piece of connected speech.

APPLICATIONS TO PRACTICE

The findings of this study, and the additional considerations just cited, suggest that pronunciation teaching programs should rank lexical stress at a medium level of importance. It is interesting that this conclusion is not dissimilar to that reached by Dalton and Seidlhofer (1994, p. 73), who, using a very different rationale, point out that lexical stress is easier to teach than intonation but has greater communicative value than the phoneme. This final section reviews a number of established procedures for teaching lexical stress and comments on their usefulness in the light of some of the findings and issues that have been discussed.

Stress Perception Exercises

English marks lexical stress in three different ways: by duration, loudness, and pitch movement. For this reason, teacher handbooks (Dalton & Seidlhofer, 1994, pp. 97–99) and pronunciation materials often recommend exercises to train the ear in distinguishing stressed from unstressed syllables. However, all the NNL participants in this study, whatever their L1, effectively used stress as a cue to word identity—suggesting that stress recognition may not be such a serious problem as is sometimes assumed. Archibald (1998, p. 184) records a much lower rate of error in perceiving stressed syllables than in producing correctly stressed words.

Weak Quality Exercises

Practice in lexical stress often includes exercises in recognizing and producing weak syllables (Dalton & Seidlhofer 1994, pp. 99–100). In the present study, both NLs and NNLs had little difficulty in identifying items where weak vowels had been replaced by full ones. This finding suggests that weak quality does not provide an important part of the

[10] It has been suggested (Cutler, Mehler, Norris & Segui, 1992) that NNLs do not develop this technique as an automatic process. However, recent research (Field, 2001) indicates that they do indeed make use of stress (or at least duration) as a word boundary marker.

access information that is used when recognizing content words and that it does not contribute importantly to intelligibility. One can conclude that practicing weak quality syllables need not be a priority for the pronunciation teacher. This precept, however, specifically concerns weak syllables that are part of larger content words; it should not be extended to function words, which were not part of the study. It seems likely that the weak quality of many functors provides an important cue that distinguishes them from content words and thus contributes importantly to the intelligibility of longer stretches of speech (Grosjean & Gee, 1987).

Presentation by Rule

One way to present English lexical stress is through a set of rules (Dalton & Seidlhofer, 1994, pp. 101–105; Kenworthy, 1987, pp. 63–65). Clearly, mastering a rule is very different from internalizing a stress pattern for a specific item. That said, this article has drawn attention to regularities of the English lexicon that may assist teaching. Students should be made aware that around 90% of content words in running speech are monosyllabic or begin with a stressed syllable (Cutler & Carter, 1987) and that the remaining 10% includes quite a large number that contain prefixes or initial syllables that resemble them.

Presentation by Vocabulary Item

Lexical stress is specific to the individual word. Clearly, therefore, the responsibility for presenting this feature falls as much on the vocabulary teacher as on the pronunciation teacher, and the oral practice of new items should include attention to their stress pattern. This is particularly important if, as postulated, the stressed syllable forms part of the access code by which the language user locates a word in his or her mental word store.

Analogy Exercises

Teachers' handbooks and pronunciation manuals greatly favor analogy exercises (Kenworthy, 1987, pp. 60–63), where students group words with similar stress patterns, find the odd word out, and so on. This approach has strong psychological validity. Corpora of *slips of the tongue* suggest that words sharing similar stress patterns are closely linked in the mind and that a word's stress pattern forms an important cue when a speaker is trying to retrieve it (Aitchison, 2003, pp. 141–142).

Applying Lexical Stress to Segmentation

As noted, an important function of lexical stress is that it enables listeners to divide stretches of continuous English speech into separate words. This segmentation technique is a critical listening skill, and it should be practiced. One might expose listeners to short stretches of authentic speech a little above their language level, then show them how they can decode the recording into words by identifying and transcribing the stressed syllables within it.

ACKNOWLEDGMENTS

For assistance with this research, many thanks are due to Martin Parrott, Conrad Heyns and the staff of International House London, and Chris George and the staff of Eurocentre, Cambridge. I also thank John Levis for perceptive comments on an earlier draft.

THE AUTHOR

John Field writes and researches on language processes, particularly on L2 listening. He has extensive experience with materials writing and teacher training in many parts of the world, including the Middle and Far East. He has recently published two books on psycholinguistics.

REFERENCES

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching, 3,* 113–122.

Aitchison, J. (2003). *Words in the mind* (3rd ed.). Oxford, England: Blackwell.

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

Anderson-Hsieh, J. R., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38,* 561–593.

Archibald, J. (1998). *Second language phonology.* Amsterdam: John Benjamins.

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

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

Creative Technology Ltd. (2003). Soundblaster Audigy 2 ZS Platinum Pro [computer software]. Singapore: Author.

Crystal, T., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America, 88,* 101–112.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201–220.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 183–196). Hillsdale, NJ: Lawrence Erlbaum.
Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24,* 381–410.
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113–121.
Dalton, C., & Seidlhofer, B. (1994). *Pronunciation.* Oxford, England: Oxford University Press.
Delattre, P. (1965). *Comparing the phonetic features of English, French, German and Spanish: An interim report.* Heidelberg, Germany: Julius Groos.
Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19,* 1–16.
Derwing, T., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48,* 393–410.
Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception. *Journal of Memory and Language, 27,* 143–165.
Field, J. (2000). *Lexical segmentation in first and foreign language listening.* Unpublished doctoral dissertation, Cambridge University, England.
Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System, 32,* 363–377.
Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech 1,* 126–152.
Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110–125.
Gimson, A. C. (1978). Towards an international pronunciation of English. In P. Strevens (Ed.), *In Honour of A. S. Hornby.* Oxford, England: Oxford University Press.
Gimson, A. C. (1994). *An introduction to the pronunciation of English* (6th ed.). London: Arnold.
Grosjean, F., & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition, 25,* 135–155.
Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38,* 201–233.
Hogg, R., &. McCully, C. B. (1987). *Metrical phonology: A coursebook.* Cambridge, England: Cambridge University Press.
Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (Ed.), *Studies in stress and accent* (pp. 37–82). Southern California Occasional Papers in Linguistics, 4. Los Angeles: University of Southern California.
Janse, E., Nooteboom, S., & Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication, 41,* 287–301.
Jenner, B. (1989). Teaching pronunciation: The common core. *Speak Out!, 4,* 2–4. Whitstable, England: IATEFL.
Jenkins, J. (2000). *The phonology of English as an international language.* Oxford, England: Oxford University Press.
Kenworthy, J. (1987). *Teaching English pronunciation.* Oxford, England: Oxford University Press.
Leech, G., Wrayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English.* Harlow, England: Longman.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71–102.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 10,* 29–63.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97. Reprinted in J. Leather (Ed.), *Phonological issues in language learning* (pp. 285–310). Oxford, England: Blackwell, 1999.
Peperkamp, S., & Dupoux, E. (1992). A typological study of stress 'deafness.' In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.
Smith, L. E., & Nelson, C. (1985). International intelligibility of English: Direction and resources. *World Englishes, 4,* 333–342.