White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk

Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/149023/

Version: Published Version

Article:
Almbark, Rana, Bouchhioua, Nadia and Hellmuth, Sam (2019) Is there an interlanguage intelligibility benefit in perception of English word stress? Loquens. ISSN: 2386-2637

https://doi.org/10.3989/loquens.2019.061

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS

University of Sheffield

UNIVERSITY of York

Loquens 6(1)
January 2019, e061
eISSN 2386-2637
https://doi.org/10.3989/loquens.2019.061

Is there an interlanguage intelligibility benefit in perception of English word stress?

Rana Almbark[1, 2], Nadia Bouchhioua[3] and Sam Hellmuth[1]

[1]University of York, United Kingdom
[2]University of Chester, United Kingdom
[3]Université de la Manouba, Tunisia
rana.alhusseinalmbark@york.ac.uk ORCID: https://orcid.org/0000-0002-4784-2497
Nadia.Bouchhioua@flm.rnu.tn ORCID: https://orcid.org/0000-0001-7602-797X
sam.hellmuth@york.ac.uk ORCID: https://orcid.org/0000-0002-0062-904X

Submitted: 05/03/2019; Accepted: 14/05/2019; Publicado online: 22/07/2019

Citation / Cómo citar este artículo: Rana Almbark, Nadia Bouchhioua and Sam Hellmuth. (2019). *Is there an interlanguage intelligibility benefit in perception of English word stress?*, 6(1), e061. https://doi.org/10.3989/loquens.2019.061

ABSTRACT: This paper asks whether there is an 'interlanguage intelligibility benefit' in perception of word-stress, as has been reported for global sentence recognition. L1 English listeners, and L2 English listeners who are L1 speakers of Arabic dialects from Jordan and Egypt, performed a binary forced-choice identification task on English near-minimal pairs (such as[ˈɒbdʒɛkt] ~ [əbˈdʒɛkt]) produced by an L1 English speaker, and two L2 English speakers from Jordan and Egypt respectively. The results show an overall advantage for L1 English listeners, which replicates the findings of an earlier study for general sentence recognition, and which is also consistent with earlier findings that L1 listeners rely more on structural knowledge than on acoustic cues in stress perception. Non-target-like L2 productions of words with final stress (which are primarily cued in L1 production by vowel reduction in the initial unstressed syllable) were less accurately recognized by L1 English listeners than by L2 listeners, but there was no evidence of a generalized advantage for L2 listeners in response to other L2 stimuli.

Keywords: interlanguage intelligibility benefit; word-stress; perception; L2 English; L1 Arabic.

RESUMEN: *¿Existe un beneficio de inteligibilidad por interlengua en la percepción del acento tónico?*–. Este trabajo pregunta si existe un "beneficio de inteligibilidad por interlengua" ('interlanguage intelligibility benefit') en la percepción del acento tónico, como se ha reportado para el reconocimiento global de oraciones. El estudio involucró a un grupo de oyentes de inglés como primera lengua (L1), y a oyentes nativos de dialectos árabes de Jordania y Egipto que utilizan inglés como segunda lengua (L2), quienes participaron como jueces perceptivos en una tarea de identificación de respuesta binaria forzada. El estímulo estuvo conformado por pares casi mínimos del inglés (por ejemplo, [ˈɒbdʒɛkt] ~ [əbˈdʒɛkt]) producidos por un hablante nativo de inglés, y por dos hablantes de inglés como L2 de dialectos de Jordania y Egipto, respectivamente. Los resultados revelan una ventaja para los oyentes nativos de inglés, lo que replica los resultados de un estudio previo sobre reconocimiento de oraciones, y también es consistente con descubrimientos anteriores que especificaron que los oyentes nativos utilizan mayormente conocimientos estructurales en la percepción del acento tonal en lugar de los marcadores de la señal acústica. Realizaciones no nativas de las palabras acentuadas en la última sílaba (que están marcadas por una reducción vocálica en la primera sílaba en inglés nativo) fueron reconocidas menos exitosamente por los oyentes nativos en inglés, pero no hubo evidencia de una ventaja generalizada para los oyentes en un segundo idioma cuando escuchan a otros hablantes del mismo primer idioma.

Palabras clave: beneficio de inteligibilidad por interlengua; acento tónico; percepción; inglés como L2; árabe como L1.

Copyright: © 2019 CSIC. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

1. INTRODUCTION

An 'interlanguage intelligibility benefit' has been reported for global sentence perception (Bent & Bradlow, 2003), whereby L2 English listeners outperform L1 English listeners in a sentence recognition task on the productions of other L2 speakers. In the present paper we explore whether a similar effect holds in the narrow domain of L2 listeners' perception of English word-stress. Specifically, we explore whether non-target-like phonetic realization of stress in L2 speakers' productions results in intelligibility issues for L1 and/or L2 listeners in a word recognition task on English stress near-minimal pairs. We use speech stimuli extracted from larger utterances elicited using a carefully controlled paradigm so that the cues to stress in the stimuli are those to word-level stress only, without any enhancement due to phrase- or sentence-level prominence. The present study thus offers a first exploration of an eventual interlanguage intelligibility benefit due to transfer of L1 patterns in the acoustic realization of stress into L2 productions. We also explore the general issue of whether non-target-like acoustic realization of word stress leads to reduced intelligibility of L2 speech, by L1 and/or L2 listeners.

We use the term 'stress' to denote word-level stress or lexical prominence, and the term 'accent' to denote phrase-level stress or post-lexical prominence. The focus of our study is word-level stress as produced and perceived by speakers of English as first (L1) and as second or additional (L2) language. We note that—to investigate stress in languages such as English and Arabic in which both stress and accent are marked (Jun, 2014)—it is necessary to control for the presence or absence of accent (Beckman & Edwards, 1994; Roettger & Gordon, 2017).

2. BACKGROUND TO THE STUDY

2.1. The correlates of stress in production

The acoustic correlates of stress have been shown to include duration, F0, overall intensity, frequency-sensitive intensity (spectral balance) and formant frequencies (F1/F2). Gordon and Roettger (2017) surveyed 110 studies on 75 languages and found that although duration was the most frequently observed cue to stress, all of these cues played a role of some kind in most of the languages surveyed. The relative strength of different cues appears to vary across languages, however.

It is widely assumed that F0 is the most prominent and consistent cue to stress in English, based on the influential early study by Fry (1955), which did not, however, examine the correlates of stress in the absence of accent. Studies which avoid the stress versus accent confound instead report duration, spectral balance and formant frequencies as the most consistent cues in English (Bouchhioua, 2016; van Heuven & Sluijter, 1996).

There has been less prior investigation of the acoustic correlates of stress in production of Arabic stress. Cross-dialectal variation in the acoustic cues to stress is likely, since cross-dialectal variation in phonological stress assignment is well established (Watson, 2011). In addition, some dialects such as Egyptian Arabic (EA) display consistent co-occurrence of stress and accent: the stressed syllable of almost all content words also carries sentence-level accent (Chahal & Hellmuth, 2015).

One of the first studies of the correlates of stress in Arabic was on Jordanian Arabic (JA), and indicated that the cues to stress in JA are duration and F1 (de Jong & Zawaydeh, 1999). In contrast, the correlates of stress reported for Tunisian Arabic are spectral balance and F1, but not duration (Bouchhioua, 2016).

In a previous study we compared the correlates of stress in JA and EA—the two dialects investigated in the present study—and found that both dialects made use of duration, intensity and F0, but not formant frequencies or spectral balance (Almbark, Bouchhioua, & Hellmuth, 2014). The only differences between JA and EA were in the degree to which cues were used: there was greater differentiation of stressed and unstressed syllables by means of duration in EA than in JA, and by means of F0 in JA than in EA. This finding for JA contrasts with that of the earlier study of JA by de Jong and Zawaydeh (1999), which did not fully control for the confound of stress and accent.

2.2. Perception of the correlates of stress

There is also cross-linguistic variation in the relative weighting of acoustic cues to stress in perception, and in the extent to which acoustic cues are relied upon compared to other factors.

Several studies have shown that listeners may rely on only a subset of the available acoustic cues in the signal. A recent study explored the perceptual behavior of English, Russian and Mandarin listeners in a forced choice identification task, in response to disyllabic pseudo-word stimuli in which F0, duration, intensity and F1/F2 of target vowels was systematically varied; vowel quality (F1/F2) had the greatest influence on the choices of listeners from all three language backgrounds, but there was variation in the relative weighting of suprasegmental cues (Chrabaszcz, Winn, Lin, & Idsardi, 2014). F0 was the next strongest cue after F1/F2 for English and Mandarin listeners, but duration and intensity were more important for Russian listeners. Similarly, Standard Mandarin listeners are influenced in their perception of stress minimal pairs, in a sequence recall task, by both duration and F0 cues; this contrasts with Taiwanese Mandarin listeners who attend primarily to F0, reflecting the lack of use of durational cues to word-level prominence asymmetries in Taiwanese Mandarin (Qin, Chien, & Tremblay, 2017). In lexical retrieval tasks, English listeners in fact rely primarily on segmental cues provided by unstressed vowel reduction: the true minimal pair 'forebear' (n.) [ˈfɔːbɛə] ~ 'forebear' (v.) [fɔːˈbɛə]—in which there are no segmental cues to stress in the form of vowel reduction in the unstressed syllable—is homophonous in perception for English listeners (Cutler, 1986).

Stress perception is also influenced by the phonological status of stress in the listener's first language (L1). French is a language which does not display word-level stress, and a sequence of studies has shown that although French listeners are able to perceive the acoustic cues to stress in an AX discrimination task, they are unable to discriminate stress minimal pairs in a sequence recall task which requires phonological encoding of those acoustic cues in lexical representations (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997); this holds even after long-term exposure to (and advanced proficiency in) Spanish, which is a language with contrastive stress (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008).

Finally, perception of stress is not influenced solely by acoustic correlates to stress and their relative weighting or phonological status. Several studies have shown that 'bottom-up' phonetic cues are used alongside 'top-down' cues such as lexico-semantic information in perception and processing of stress (Cole, Mo, & Hasegawa-Johnson, 2010; Eriksson, Thunberg, & Traunmüller, 2001). Mattys, White, and Melhorn (2005) argue that English listeners rely on different types of cues in a word segmentation task, with cues forming a hierarchy: fine-grained phonetic cues to stress are argued to be lower in the hierarchy than lexical and semantic cues, because phonetic cues are only relied on when performing the task in adverse listening conditions. This may be one strategy which allows listeners to use 'perceptual normalization' to recover the hypothesized intended form from non-target-like realizations (Ohala, 1993). In contrast, L2 listeners show less reliance than L1 listeners on 'top-down' structural or lexical information in a word-by-word prominence rating task; instead, L2 listeners' ratings more closely reflected differences in the relative strength of acoustic phonetic cues (Wagner, 2005).

2.3. The interlanguage intelligibility benefit

The term 'interlanguage' describes patterns of language use, displayed by second language learners, which fall somewhere between the grammar of the native language and the target language being acquired (Selinker, 1972).

The concept of an interlanguage speech intelligibility benefit was proposed by Bent and Bradlow (2003) to explain their findings in a sentence recognition task performed on L1 and L2 English speech samples, by L1 English listeners in comparison to L2 English listeners whose L1 varied. For native English listeners, the native English speech was more intelligible (more keywords accurately recognized) than the L2 English speech; however, for the L2 English listeners, the L1 English and L2 English speech were equally intelligible, regardless of whether the L2 English listener's L1 background matched that of the L2 English speaker they were listening to.

The two main groups of L2 English listeners in the Bent and Bradlow (2003) study were L1 speakers of Chinese and Korean. Stibbard and Lee (2006) replicated the same study design with L2 English speakers/listeners from more typologically diverse L1 backgrounds, however, and obtained a more nuanced result. They explored the perceptual behavior of L2 English speakers from Saudi Arabia or Korea, at two proficiency levels in English (low and high). In their study, the L1 English listener group showed higher recognition rates than any of the L2 listener groups, but high proficiency L2 English samples were equally well recognized as L1 English samples by both L1 and L2 listeners. The main finding of the replication study was that low proficiency was highly correlated with low intelligibility, as might be expected, but also that there was a matched interlanguage speech intelligibility benefit: low proficiency L2 English speech was better recognized by L2 listeners from the same L1 background as the speaker in the L2 English sample.

In this study we explore whether there is an interlanguage speech intelligibility benefit in respect of L1 versus L2 realization of the phonetic cues to word-level stress.

2.4. The present study

The main research question of the paper is to determine whether there is an interlanguage intelligibility benefit in perception of English word stress. We use stimuli that were elicited using a paradigm designed to elicit English stress near-minimal pairs in a context in which the target word is realized without a phrase-level accent, thus focusing on listeners' ability to make use of the phonetic cues to stress in the absence of cues to accent. Since vowel reduction is the primary cue to word stress for native English listeners (as noted in 2.2 above), it was important to use stimuli in which vowel reduction could appear, to determine whether failure to produce target words with appropriate vowel reduction reduces intelligibility, and perhaps differentially so for native versus non-native listeners. We therefore used near-minimal pairs in which vowel reduction in the unstressed syllable provides a segmental cue to stress alongside suprasegmental cues such as duration and intensity. The stimuli were produced by an L1 English native speaker (NE) and two L2 English non-native speakers (L2) from Jordan and Egypt, respectively. The listeners in a forced-choice identification task are L1 English listeners (NE) and L2 English listeners from Jordan and Egypt (L2). The over-arching research question stated in the title of this paper thus breaks down into three sub-questions, which we address in the present study by exploring the interaction of listener language and stimulus language in a single study with a crossed factor design:

1. Do NE listeners identify the position of stress in the productions of a NE speaker more accurately than in those of L2 speakers?
2. Do L2 listeners identify the position of stress in the productions of L2 speakers more accurately than in those of a NE speaker?
3. Do L2 listeners identify the position of stress in the productions of an L2 speaker from their own L1 dialect background more accurately than in those of an L2 speaker from a different L1 dialect background?

Based on Bent and Bradlow's (2003) findings, we would predict an advantage for NE listeners when listening to NE productions, but no advantage for L2 listeners when listening to other L2 listeners (from any background). Based on Stibbard and Lee's (2006) findings, however, we predict an overall advantage for NE listeners, but a possible advantage for L2 listeners when listening to L2 listeners. Our interpretation of the results will also consider whether there are differences between NE and L2 listeners in reliance on 'bottom-up' phonetic cues versus 'top-down' structure-based expectations, by examining possible transfer effects which reflect the different structural properties of stress assignment in listeners' L1.

3. METHODS

3.1. Materials

Stimuli which contrast in the position of stress were elicited using the nine English disyllabic near-minimal pairs, listed in Table 1, following Bouchhioua (2008, 2016).

Three further pairs (*combine*, *pervert*, and *project*) were recorded but later excluded from the study, as stress was frequently misplaced due to unfamiliarity with the word in one or both stress positions. Six target-like tokens of the word *project* (two from each speaker) were used for the training phase of the experiment as outlined further below.

The intended accent status of the target word was varied by using a carrier phrase that either attracts focus to the target word [+accent] or diverts focus away from it [−accent], again following Bouchhioua (2008, 2016), as shown in Table 2. The target word was always elicited in a carrier phrase: 'say ___ again'. To attract accent onto the target word, a semantically related word preceded the target word in the same carrier phrase. To divert focus away from the target word, two preceding sentences are used to ensure that the target word appears in post-focal position (after the contrastively focused verb 'SAY') and is interpreted as old information due to being repeated from the immediately preceding discourse (Cruttenden, 2006; Ladd, 2008). Each sentence ~ context combination was read aloud once; sentences were presented to participants in pseudo-random order on a printed sheet.

The experimental stimuli for the present study were extracted from target-like tokens (as judged to consensus by the first and third authors) produced in *−accent* condition, as in (1), to investigate the extent to which listeners were able to detect phonetic cues to stress produced by the speakers, in the absence of any additional cues to accent.

(1)

stress on first syllable:	SAY ˈsʌbdʒɛkt again.
stress on second syllable:	SAY səbˈdʒɛkt again

Table 1: English near-minimal pairs, with stress on the first or second syllable.

stress on first syllable		stress on second syllable	
ˈsʌbdʒɛkt	subject (n.)	səbˈdʒɛkt	subject (v.)
ˈɹɛkɔːd	record (n.)	ɹɪˈkɔːd	record (v.)
ˈkɒntɹæst	contrast (n.)	kənˈtɹæst	contrast (v.)
ˈdaɪdʒɛst	digest (n.)	dɪˈdʒɛst	digest (v.)
ˈkɒntɹækt	contract (n.)	kənˈtɹækt	contract (v.)
ˈpɜːmɪt	permit (n.)	pəˈmɪt	permit (v.)
ˈɒbdʒɛkt	object (n.)	əbˈdʒɛkt	object (v.)
ˈkɒntɛnt	content (n.)	kənˈtɛnt	content (adj.)
ˈkɒndʌkt	conduct (n.)	kənˈdʌkt	conduct (v.)

Table 2: Target word (in **bold**) placed in carrier phrases to vary ±accent status. CAPITALS denote expected position of sentence accents under focus.

+accent	Say topic again. Say SUBJECT again.
−accent	The subject is a grammatical category. WRITE subject again. SAY **subject** again. ←

The stimuli for the perception experiment were produced by three male speakers, from: Cairo, Egypt (EA); Amman, Jordan (JA); UK (native speaker of British English, NE). The speakers were aged 26, 20, and 39 years, respectively. The Arabic speakers had learned English at school for 12 years but had never resided in an English-speaking country; they were selected from participants in an earlier production study (Almbark *et al.*, 2014). Recordings were made in Cairo, Amman and York, respectively. Recordings were made in .wav format at 44.1 KHz 16 bit, on a Marantz PMD660 with external Shure SM10 headset microphone.

The results of acoustic analysis of the selected stimuli for duration, F0, intensity, F1/F2 and two measures of spectral tilt (H1.H2 or H1.A3), comparing properties of the vowel in the initial syllable (only), in stressed and unstressed condition, are illustrated in Figures 1–2. We used a normalized vowel duration measure to control for inter-speaker variation in speech rate, by calculating vowel duration as a proportion of the whole word. The acoustic properties of the stimuli were explored in a series of linear mixed models (LMM) using *lme4* (Bates, Maechler, Bolker, & Walker, 2015) in *R* (Core Team, 2014), with each acoustic measure in turn as dependent variable, *speaker* (EA ~ JA ~ NE) and *stress* (stressed ~ unstressed) and their interaction as fixed factors, and a random intercept for *item*.

The acoustic analysis shows that F0 differentiates stressed and unstressed syllables in the L2 English productions of the JA speaker, but not in those of the EA or NE speakers. Similarly, although intensity is somewhat

Figure 1: Median and interquartile range for values of (from left to right) maximum F0, peak intensity, normalized vowel duration and two measures of spectral emphasis H1–H2 and H1A3, in the first vowel of experimental stimuli produced by the native speaker of Egyptian Arabic (EA; top row), Jordanian Arabic (JA; middle row), and English native speaker (NE; bottom row), grouped by stress condition (whether the vowel in which measurements was taken was stressed or unstressed).

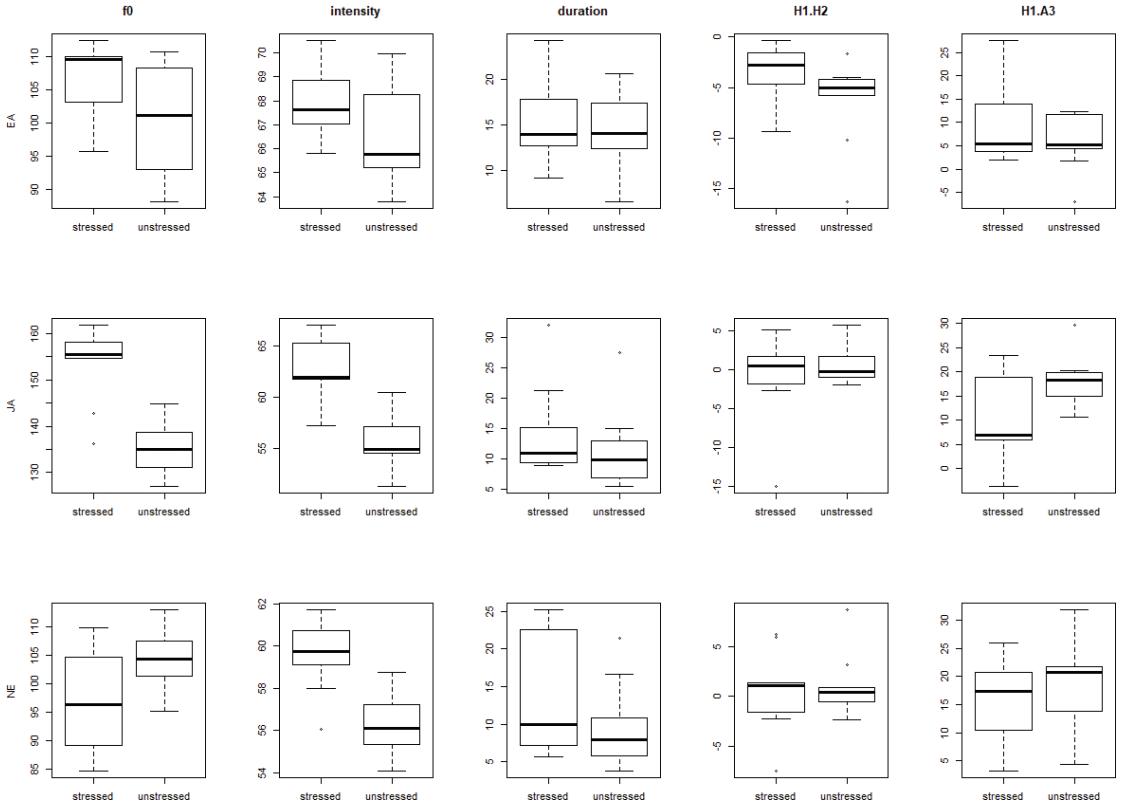
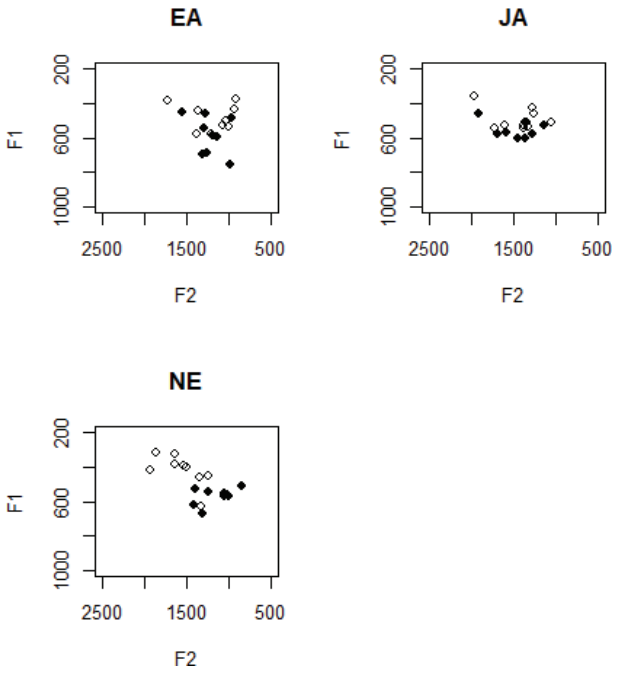


Figure 2: F1/F2 plot of the first vowel in experimental stimuli produced by the native speaker of Egyptian Arabic (EA), Jordanian Arabic (JA) and English (NE), where the vowel is stressed (black dots) or unstressed (white dots).



higher in stressed syllables than unstressed syllables for all three speakers, including the EA speaker (for whom intensity is the strongest cue to stress on average), nevertheless it is only in the JA speaker's production that this difference is significant. In contrast, neither vowel duration nor spectral tilt (H1.H2 or H1.A3) is used to differentiate stressed and unstressed syllables by any of the speakers. Finally, both F1 and F2 differentiate stressed and unstressed syllables to a significant extent in the NE speaker's productions, but not in the productions of the EA speaker and JA speaker.

The differences among the three speakers in the observed cues to stress in the experimental stimuli match the generalizations reported for the full set of speakers who participated in the study from which the stimuli were extracted (Almbark *et al.*, 2014), which also reports the phonetic realization of stress in L1 EA and JA by the same speakers.

3.2. Participants

Participants were recruited by email invitation among the friends and family of graduate students of linguistics from Egypt and Jordan, and among students at the University of York. A total of 42 listeners meeting our inclusion criteria (by native language/dialect, excluding

early bilinguals) completed the online perception experiment on a voluntary basis. From these a balanced subset of 36 was selected at random to yield three listener groups by native language: EA, JA or English (NE), with six male and six female listeners in each group. The Arabic-speaking listeners had all studied English for at least 12 years; six had English medium schooling (two EA, four JA); one JA listener was in the UK at the time of taking test.

3.3. Procedure

The experiment was run using an online survey tool (SurveyGizmo, 2019). Participants first read an information sheet and provided their informed consent to participate; they then completed a questionnaire about age, sex, native language and dialect, and, for L2 listeners, number of years of study of English.

Participants were familiarized with the test paradigm in a training phase; a selection of English stimuli were presented, which differed in stress position as in the main test, using the target word 'project' [ˈpɹɒdʒɛkt] ~ [pɹəˈdʒɛkt]. Participants were asked to answer the following question for each word they heard: "Was it PROject (first syllable) or proJECT (second syllable)?". Feedback was given as to whether the provided answer was correct or incorrect.

After the training phase, in the first test phase the 36 sound files produced by the two L2 English speakers (9 target words × 2 stress conditions × 2 speakers = 36) were presented in randomized order. Each sound file was shown on a separate page with the question "Is it ___ (first syllable) or ___ (second syllable)?" and two answers (e.g., "SUBject with stress on the first syllable" or "subJECT with stress on the second syllable") to choose from, in a binary forced choice. Then, in the second test phase, the 18 sound files produced by the L1 English speaker were presented, following the same procedure as for the first test phase.

We presented all L2 speech in one block, then all L1 speech in a separate block, to restrict the listeners' task to word recognition. Randomisation of tokens extracted from L1 and L2 speech in one block might have drawn listeners' attention to evaluation of the degree of foreign accent rather than the intelligibility (i.e., recognition) of the utterances as intended.

3.4. Analysis

Each response was coded for *accuracy*: responses which matched the intended form of the word as elicited were coded as correct, otherwise as incorrect. Results were explored using binomial generalized linear mixed models (GLMM) using *lme4* (Bates *et al.*, 2015) in *R* (Core Team, 2014), with *accuracy* as the dependent variable, using likelihood ratio tests to identify the best fit model. The predictions of the model were extracted using *lsmeans* (Lenth, 2016) and plots were produced using *ggplot2* (Wickham, 2009).

4. RESULTS

Figure 3 shows accuracy rates for the three groups of listeners, grouped by stimulus language and elicited position of stress. Accuracy rates are above chance for most participants (where chance would equate to a score of 4 or 5, in a binary forced choice task with a maximum score of 9). Accuracy is above chance for English listeners in response to all stimuli produced by the NE speaker, and there is a ceiling effect for English listeners in response to stimuli elicited with initial stress. Visually, it appears that English listeners are somewhat more accurate than EA listeners, who are in turn somewhat more accurate than JA listeners, but that there is little effect of stimulus language for the Arabic listeners.

However, any variation across listener groups is clearly mediated by variation within listener groups that reflects the elicited position of stress in the word: EA listeners are less accurate at identifying words produced by the English speaker with initial stress; English listeners, in turn, are less accurate at identifying words produced by the Egyptian speaker with final stress. In contrast, accuracy rates of JA listeners show largely overlapping distributions by both position of stress and by stimulus language.

These effects were explored in a series of GLMM models; the best fit model includes fixed factors for stress condition (*stress*), listener language *(listlang)* and stimulus language *(stimlang)*, and all interactions among these three factors, with random intercepts for *participant* and *item*. Separate models were run including the control factors *age*, *sex*, and *device* (encoding participants' use of earphones versus external loudspeaker to take the test), but none of these factors improved model fit. The best fit model summary is reported in Table 3. The reference levels for the fixed factors were 'initial' (for *stress*) and 'EA' (for *listlang* and *stimlang*); the model was re-run with 'JA' as reference level to obtain

Figure 3: Median (bold vertical line within bars) and interquartile range (bar size) of the count of accurate responses for each individual participant, grouped by listener language, stimulus language, and elicited position of stress. EA = Egyptian Arabic; JA = Jordanian Arabic.

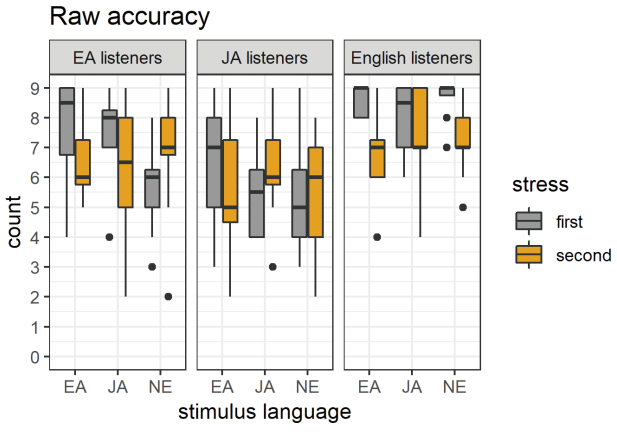


Table 3: Summary of the best fit GLMM [accuracy ~ listlang * stimlang * stress + (1 | item) + (1 | participant)].

Fixed effects	Estimate (log odds)	*SE*	*z*	*p*
intercept	1.88724	0.34165	5.524	< .000
stimlangJA	-0.07544	0.38553	-0.196	0.8448
stimlangNE	-1.30793	0.34590	-3.781	0.0001
listlangJA	-0.79566	0.41624	-1.912	0.0559
listlangNE	1.57067	0.61761	2.543	0.0109*
stresssecond	-0.84735	0.39831	-2.127	0.0333
stimlangJA:listlangJA	-0.51935	0.49107	-1.058	0.2902
stimlangNE:listlangJA	0.58733	0.45942	1.278	0.2011
stimlangJA:listlangNE	-1.03615	0.71293	-1.453	0.1461
stimlangNE:listlangNE	1.30792	0.79567	1.644	0.1002
stimlangJA:stresssecond	-0.07017	0.49480	-0.142	0.8872
stimlangNE:stresssecond	1.57107	0.47366	3.317	0.0009***
listlangJA:stresssecond	0.25170	0.46696	0.539	0.5898
listlangNE:stresssecond	-1.40917	0.65985	-2.136	0.0327*
stimlangJA:listlangJA:stresssecond	1.01822	0.65254	1.560	0.1186
stimlangNE:listlangJA:stresssecond	-0.85047	0.63254	-1.345	0.1787
stimlangJA:listlangNE:stresssecond	1.59293	0.84933	1.876	0.0607
stimlangNE:listlangNE:stresssecond	-1.28632	0.92208	-1.395	0.1630

pairwise comparisons (which are reported where relevant in the text).

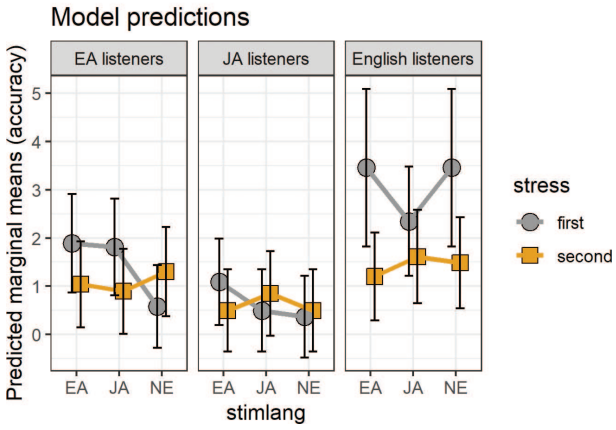
For the dependent variable the reference level in all models was 'incorrect'; the models thus predict the log odds of improved accuracy resulting from a change in *stress* or *stimlang* or *listlang* condition or a combination of these. The predicted marginal means of the model, and 95% confidence intervals around them, are illustrated in Figure 4; this plot visualizes the significant effects predicted by the model (overlapping confidence intervals indicate an effect which is not significant).

The best fit model shows no significant three-way interactions and no main effect of stimulus language or stress position. There were no significant interactions between listener language and stimulus language; it is this type of interaction that would indicate an interlanguage intelligibility benefit.

There is a main effect of listener language: English listeners are much more accurate than JA listeners ($z(1924) = 3.967$; $p < .000$) and also somewhat more accurate than EA listeners ($z(1924) = 2.543$; $p = .010$), regardless of speaker language and stress position. This matches the pattern observed for English listeners by Stibbard and Lee (2006).

There is a significant interaction between stress and listener language: English listeners were less accurate at identifying words with final stress across the board, regardless of stimulus language ($z(1924) = -2.136$; $p = .0327$). There was also a significant interaction between stress and stimulus language: words with final stress were less accurately identified by all listeners when produced by the EA speaker, than either the NE speaker ($z(1924) = 3.317$; $p = .0009$) or the JA speaker ($z(1924) = 2.227$; $p = .0259$). We explore these interactions with stress position in the general discussion below.

Figure 4: Predicted marginal means (and 95% CI) for the best fit binomial GLMM by listener language, stimulus language (*stimlang*), and position of stress. EA = Egyptian Arabic; JA = Jordanian Arabic; NE = English native speaker.



5. DISCUSSION

Our specific research question was to explore a possible interlanguage intelligibility benefit in perception of English word stress; that is, to test the hypothesis that L2 listeners will more accurately interpret English word stress when produced by other L2 speakers. We found no evidence to support this hypothesis in this study, as

there are no significant interactions between any levels of listener language and stimulus language in our data. This also rules out the type of interlanguage intelligibility benefit found by Stibbard and Lee (2006), where L2 listeners perform better when listening to speakers from the same L2 background: the distribution of accuracy rates for EA listeners in response to EA stimuli overlaps with that observed in response to JA stimuli (and likewise, the distribution of accuracy rates for JA listeners in response to JA stimuli overlaps with that in response to EA stimuli). We thus find no evidence for an interlanguage intelligibility benefit based on phonetic realization of stress.

Our results replicate the finding of Stibbard and Lee (2006) who also found that English listeners performed better in a sentence recognition task across the board, in comparison to L2 listeners. Our study extends this finding to include recognition of lexical items differentiated solely by stress, in response to stimuli which bear cues to stress only, without any additional enhancement in cues due to phrase-level accent. We attribute this finding to the ability of L1 listeners to make use of 'top-down' structural and/or lexico-semantic cues in perception of stress; in the present study this could be because the native English listeners are more familiar with the lexical items used as stimuli than L2 learners are. The lower accuracy of the L2 listeners in our results, across the board, mirrors the findings of other studies which showed that L2 listeners are less reliant on 'top-down' cues; in the present study this may be a direct effect of reduced familiarity with some of the lexical items, and/or reduced of awareness of the existence of stress near-minimal pairs in English. These competing explanations could be explored in future research by using pseudoword stimuli or by controlling for L2 learners' vocabulary size.

The study shows two significant interactions of listener/speaker language with the position of stress. The first of these is that NE listeners displayed lower accuracy in response to words with final stress, regardless of speaker language. The expected NE realization in these words has vowel reduction in the first syllable, to schwa [ə] in 7 out of 9 of our stimuli, and to [ɪ] in the other two cases (see Table 1). Vowel reduction is in fact the primary cue to stress for English listeners (Cutler, 1986), so this result suggests that the reduced vowel reduction in the stimuli produced by the two L2 speakers (illustrated in Figure 2) may indeed have contributed to lower intelligibility of their productions by the NE listeners. The second significant interaction with stress position was that all listeners were less accurate in their interpretation of the EA speaker's productions of words with final stress. We attribute this reduced accuracy to the reduced differentiation of stressed and unstressed syllables in the productions of the EA speaker (see Figure 1); this lack of differentiation may in turn result from the previously reported conflation of word- and phrase-level stress in this dialect (Hellmuth, 2007). Taken together we interpret these interactions as evidence that non-target-like phonetic realization of stress can result in lower intelligibility of L2 speakers' productions for both L1 and L2 listeners in certain contexts.

6. CONCLUSION

The aim of this paper was to explore a possible interlanguage intelligibility benefit for L2 listeners in perception of stress, due to potential transfer of L1 patterns of phonetic realization of stress into L2 productions. The results did not show any interlanguage intelligibility benefit but did confirm the previous finding of an overall advantage for L1 English listeners in lexical recognition tasks, which we attribute to the L1 listeners' ability to make use of top-down lexical knowledge in perception of stress. This strategy supports accurate recognition in the face of non-target-like phonetic cues to stress encountered in L2 English productions, but we show that non-target-like cues can result in reduced intelligibility of L2 speakers when the primary cue expected by L1 listeners (here, vowel reduction) is the same cue that the L2 speaker fails to produce to a target-like extent.

REFERENCES

Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014). Acquiring the phonetics and phonology of English word stress: Comparing learners from different L1 backgrounds. *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics, 5*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (Ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III* (pp. 7–33). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.002

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bouchhioua, N. (2008). *The acoustic correlates of stress and accent in Tunisian Arabic: A comparative study with English* [Unpublished PhD dissertation]. University of Carthage, Tunisia.

Bouchhioua, N. (2016). Typological variation in the phonetic realization of lexical and phrasal stress: Southern British English vs. Tunisian Arabic. *Loquens, 3*(2), e034. https://doi.org/10.3989/loquens.2016.034

Chahal, D., & Hellmuth, S. (2015). Comparing the intonational phonology of Lebanese and Egyptian Arabic. In S.-A. Jun (Ed.), *Prosodic typology. Volume 2* (pp. 365–404). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.003.0013

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1*(2), 425–452. https://doi.org/10.1515/labphon.2010.022

Core Team, R. (2014). *R: A language and environment for statistical computing*. Vienna: Austria. Retrieved from http://www.R-project.org/

Cruttenden, A. (2006). The de-accenting of old information: A cognitive universal. In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic organization of discourse in the languages of Europe* (pp. 311–355). Berlin: Mouton de Gruyter.

Cutler, A. (1986). Forebear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201–220. https://doi.org/10.1177/002383098602900302

de Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics, 27*(1), 3–22. https://doi.org/10.1006/jpho.1998.0088

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language, 36*, 406–421. https://doi.org/10.1006/jmla.1996.2500

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition: International Journal of Cognitive Science, 106*, 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

Eriksson, A., Thunberg, G. C., & Traunmüller, H. (2001). Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. *EUROSPEECH-2001*, 399–402.

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*(4), 765–768. https://doi.org/10.1121/1.1908022

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard, 3*(1). https://doi.org/10.1515/lingvan-2017-0007

Hellmuth, S. (2007). The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review, 24*(2–3), 289–314. https://doi.org/10.1515/TLR.2007.011

Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology Volume II.* (pp. 520–539). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.001.0001

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple segentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500. https://doi.org/10.1037/0096-3445.134.4.477

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.

Qin, Z., Chien, Y.-F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics, 38*(3), 541–570. https://doi.org/10.1017/S0142716416000321

Roettger, T., & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistics Vanguard, 3*(1). https://doi.org/10.1515/lingvan-2017-0006

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*(1–4), 209–232. https://doi.org/10.1515/iral.1972.10.1-4.209

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America, 120*(1), 433–442. https://doi.org/10.1121/1.2203595

SurveyGizmo. (2019). *SurveyGizmo | Enterprise Online Survey Software & Tools*. Retrieved from https://www.surveygizmo.com/

van Heuven, V., & Sluijter, A. (1996). Notes on the phonetics of word prosody. In R. Goedmans, H. van der Hulst, & E. Visch (Eds.), *Stress patterns of the world. Part 1: Background.* (pp. 233–269). The Hague: Holland Academic Graphics.

Wagner, P. (2005). Great expectations-introspective vs. perceptual prominence ratings and their acoustic correlates. *INTERSPEECH-2005*, pp. 2381–2384.

Watson, J. C. E. (2011). Word stress in Arabic. In M. Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology. Volume 5* (pp. 2990–3018). Oxford: Blackwell. https://doi.org/10.1002/9781444335262.wbctp0124

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer Science & Business Media.