Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 14 (4). 2024. 661-686. Published online: 24.09.2024
*https://doi.org/10.14746/sslt.38361*
http://pressto.amu.edu.pl/index.php/sslt

# *The effects of implicit corrective feedback on production of lexical stress in L2 English*

Özgür Parlak ✉
American University of Sharjah, United Arab Emirates
https://orcid.org/0000-0003-3174-8700
*oparlak@aus.edu*

Abstract

The interactionist approach to second language acquisition has yielded a plethora of studies confirming the positive impact of interaction and corrective feedback on second language (L2) development. Nevertheless, only a few studies have attempted to investigate the development of L2 prosody using the interactionist approach. The current study contributes to this line of research by investigating the relationship between recasts and the production of primary stress in L2 English. Following a pretest-posttest design, 68 L1 Arabic speakers were randomly assigned to control and intervention groups. The pre- and posttest comprised sentence-completion and information-exchange tasks, whereas the intervention was a role-play task that dyads carried out with the researcher. The intervention group received a recast upon producing target words with misplaced primary stress, whereas the control group did not receive any corrective feedback. The results of acoustic analyses, which focused on syllable duration, intensity, and pitch, indicated a positive relationship between recasts and development of primary stress placement. The results were also supported by expert listener judgments. The findings suggest that interaction and implicit corrective feedback play a positive role in the development of lexical stress.

*Keywords*: interaction; corrective feedback; recasts; L2 pronunciation; lexical stress

## 1. Introduction

Emanating from the interaction hypothesis (Long, 1996), the interactionist approach (Gass & Mackey, 2015) to second language acquisition (SLA) posits that communicative interaction aids second language (L2) development as it provides opportunities for receiving input, producing output, and negotiating for meaning. Input triggers language processing, which is necessary for comprehension of new forms and the consolidation of previously encountered forms (Gass, 1988; Long, 1990). On the other hand, output formulation requires the utilization of linguistic resources to create meaning (Swain, 2005). Errors in learners' output may initiate negotiation for meaning and the provision of corrective feedback, which helps learners notice the linguistic gap between their non-target-like output and the target form (Schmidt, 2001). The effectiveness of corrective feedback has been researched extensively, with several meta-analyses and reviews highlighting its positive impact on L2 development (e.g., Brown, 2016; Goo & Mackey, 2013; Li, 2010; Ziegler, 2016). However, corrective feedback studies conducted over the past three decades have primarily focused on the development of morphosyntactic features with a relatively small number of studies focusing on L2 phonology (Mackey et al., 2012). Even fewer studies have investigated the development of suprasegmental features using the interactionist approach. The current study contributes to this particular area of investigation by reporting the acoustic changes that occurred in first language (L1) Arabic speakers' primary stress production in English subsequent to receiving recasts during a series of meaning-focused tasks.

## 2. Background

### 2.1. Corrective feedback and L2 phonology

Empirical evidence shows that implicit corrective feedback, mainly recasts, can have a positive impact on L2 phonological development during form-focused instruction or in a laboratory setting (Bryfonski & Ma, 2020; Lee & Lyster, 2016; Parlak & Ziegler, 2017; Saito & Lyster, 2012; Saito & Wu, 2014). For example, Saito's (2013) research on segmental features indicated that recasts provided during form-focused instruction are likely to promote perceptual and productive development of the approximant /ɹ/. In another study, Lee and Lyster (2016) found that recasts provided during form-focused instruction facilitated perceptual development of the /i/-/ɪ/ phonemic contrast. As for suprasegmental features, Saito and Wu (2014) reported that form-focused instruction with recast provision improved perception of Mandarin tones only under the trained vocabulary

condition, whereas form-focused instruction without recasts led to significant gains under all conditions. The authors concluded that recasts promote lexical learning rather than a general development of tone perception. Bryfonski and Ma (2020) also investigated the development of Mandarin tones by comparing the effectiveness of metalinguistic feedback to that of recasts. Their findings showed that receiving recasts was associated with higher levels of productive development compared to receiving metalinguistic feedback. There was no statistical difference between the two feedback moves in terms of perceptual development. Finally, Parlak and Ziegler (2017) conducted a study comparing the effects of recasts on the development of lexical stress in L2 English in face-to-face and synchronous computer-mediated conditions by focusing on four-syllable and three-syllable words. Although they did not find significant gains at the group level, analysis of three-syllable words showed that the face-to-face group produced target syllables with statistically longer duration on the posttest. As duration is a robust correlate of lexical stress (Gordon & Roettger, 2017; Kochanski et al., 2005), the authors concluded that recasts potentially have a positive impact on the development of lexical stress and noted that complexity and readiness may mediate the usefulness of recasts. In short, the relatively recent yet growing number of interactionist studies provide preliminary support for the potential usefulness of corrective feedback, in particular recasts, on L2 phonological development.

## 2.2. The communicative value of lexical stress

There are surprisingly few corrective feedback studies targeting suprasegmental features despite their role in effective verbal communication. One key suprasegmental feature in English is lexical stress, which is a building block of word formation realized at the syllable level. Stressed syllables are produced with more vocal energy, which causes them to be perceived as more prominent relative to other syllables. Polysyllabic words in English have one syllable that carries primary stress. However, the placement of primary stress in English is not always predictable, and that can be challenging for learners. In some cases, stress is paradigmatic and its placement follows a certain pattern depending on the part of speech (e.g., noun-verb homographs as in PROject vs. proJECT) or the attachment of a stress-shifting derivational suffix (e.g., the suffix *-ic* causing primary stress to move to the preceding syllable as in HARmony vs. harMOnic). On the other hand, the assignment of primary stress in English can also be lexical (e.g., PAnama vs. baNAna), which makes prediction difficult. As for Arabic, the L1 of participants in the current study, lexical stress placement is determined by syllable weight (Hellmuth, 2013), and, therefore, it is more predictable. Generally,

super-heavy ultimate (CVCC or CVVC) or heavy penultimate (CVC or CVV) syllables carry stress with only occasional differences across different dialects of Arabic (Watson, 2007). In other words, lexical stress placement follows a more systematic pattern in Arabic than English. As far as the realization of stress goes, the acoustic correlates of primary stress are mainly duration, intensity, and pitch in both English and Arabic (Almbark et al., 2014).

Lexical stress has high communicative value in English and inaccurate stress could impact verbal interaction negatively (Lewis & Deterding, 2018). For example, Kang et al. (2010) observed suprasegmental features to collectively account for 52% of variance in oral proficiency ratings and 50% of the variance in comprehensibility judgments. The authors also found a direct relationship between correct lexical stress placement and higher comprehensibility judgments. In another study by Saito et al. (2016), lexical stress was found to determine oral ability for all proficiency levels, while segmental accuracy was found to be a predictor for oral ability only at the advanced level. Research has also shown that non-target-like stress placement is associated with lower intelligibility (Field, 2005), comprehensibility, and higher accentedness ratings (Isaacs & Trofimovich, 2012; Saito et al., 2016). As highlighted by these studies, correct placement of lexical stress is essential for achieving higher oral proficiency and effective verbal communication.

## 2.3. The choice of recasts

Recasts are a form of reactive corrective feedback which reformulate learners' erroneous production into a more acceptable form in the target language. Recasts commonly occur in natural (Braidi, 2002) as well as instructional settings (Sheen, 2004). Research to date has identified several characteristics which make recasts an ideal feedback move when addressing pronunciation errors. First and foremost, recasts are minimally disruptive allowing interaction to flow naturally (Gass & Mackey, 2006). Also, they signal an error while simultaneously providing a model; in other words, they provide both negative and positive evidence (Leeman, 2003). These two features make recasts particularly useful when addressing pronunciation errors as learners may feel frustrated when interrupted with a more explicit form of corrective feedback (Bryfonski & Ma, 2020) or when they are not provided with a model that would help them improve their pronunciation. Finally, the contingency of recasts to the erroneous production increases the salience of recasts and gives learners a chance to juxtapose their production with the recast (Long, 2007). This is probably one of the reasons why learners readily notice the intent of recasts when the target is a pronunciation error (Carpenter et al., 2006; Mackey et al., 2000; Scheuer & Horgues, 2019). As a form of

implicit feedback that is non-intrusive but at the same time one that can be made relatively more explicit with the use of additional stress, intonation, as well as contingency, recasts are an ideal form of corrective feedback when addressing lexical stress errors.

### 2.4. The current study

The current study investigated the impact of recasts on the production of primary stress in L2 English. The study took place in a laboratory setting and followed a pretest-posttest design with task-based intervention. The participants, who were L1 speakers of Arabic, carried out a series of pre-task activities, two information-exchange tasks and an interview role-play task during which the intervention group received recasts. In order to gain a finer understanding of the physical changes that recasts can potentially trigger, the study adopted a combination of acoustic and auditory measures of development to address the following research questions:

RQ1: How does recast provision targeting learners' primary stress errors impact acoustic realization of primary stress?
RQ2: To what extent is providing recasts on learners' primary stress errors associated with target-like production based on expert listener judgments?

## 3. Method

### 3.1. Participants

Participants were 74 learners enrolled in an Intensive English Program at a university in the Arabian Gulf. They were recruited through classroom announcements and given a gift voucher for a cup of coffee and a donut for their participation. The researcher had no previous relationship with the participants. Six data sets were removed because one belonged to an L1 Hindi speaker and five were unusable due to problems with the recordings. The remaining data were from 68 participants who were speakers of eastern Arabic dialects (Mustafawi, 2018) with a mean age of 18 ($SD$ = 0.86). According to the self-reported TOEFL ($N$ = 19) and IELTS ($N$ = 49) scores, participants were mainly upper-intermediate level learners. For ease of comparison, TOEFL scores were converted to IELTS scores following the conversion tables provided by Educational Testing Service (2010). Apart from one participant who reported difficulty with pronouncing the

phoneme /ɹ/, participants did not disclose any hearing- or speech-related problems (see Table 1 for an overview).

Table 1 Participant demographics

| | | Control group | Intervention group |
|---|---|---|---|
| Age (years) | | 17.78 (*SD* = 0.75) | 18.19 (*SD* = 0.92) |
| Sex | Male | 21 | 21 |
| | Female | 11 | 15 |
| Arabic dialect | Gulf | 15 | 20 |
| | Levantine | 10 | 11 |
| | Egyptian | 5 | 5 |
| | Sudanese | 2 | 0 |
| IELTS scores | 5.5-6.0 | 31 | 32 |
| | 6.5-7.0 | 1 | 4 |

## 3.2. Materials

The target words were three-syllable words with primary stress on the penultimate syllable (see Table 2). They were words that the same target population, but different participants, struggled with in terms of stress placement during the pilot study (see Parlak & Ziegler, 2017). The topic of the tasks and sentence-completion activities, which are described below, was also considered when choosing the target words so that they allow for the formulation of meaningful carrier sentences.

Table 2 Target words

| | | | | | |
|---|---|---|---|---|---|
| Version A | *dynamic* | *develop* | *revision* | *consider* | *suspicion* |
| | *confusion* | *perception* | *responsive* | *contention* | *consistent* |
| Version B | *position* | *duration* | *diminish* | *formation* | *condition* |
| | *compassion* | *companion* | *consensus* | *convention* | *persistent* |

The study followed a laboratory design using two sentence-completion activities, two information-exchange tasks, and an interview role-play task. The main goal was to get participants to brainstorm about the ideal practices in a language classroom and then interview a candidate (the researcher) for a language teaching position. A total of three handouts were used to facilitate interaction. Handout 1 was a sentence-completion activity. Handout 2 was a list of interview questions. Finally, Handout 3 was another sentence-completion activity. Each handout comprised 10 carrier sentences with the target words embedded in sentence-medial position. Also, each of the three handouts had two versions: Version A was intended for Student A and Version B was intended for Student B in each dyad. The two versions had different sets of carrier sentences and target words

in order to prevent priming of a target word between the two participants. Example 1 shows the carrier sentences for the target words *develop* and *diminish*, with ellipses indicating the missing parts that participants were asked to complete during the sentence-completion activities.

Example 1 Carrier sentence 10 taken from handouts 1a, 1b, 2a, 2b, 3a, and 3b.

| | Version A | Version B |
|---|---|---|
| Handout 1 | When teachers *develop* a good relationship with their students. . . | In order to *diminish* issues related to language anxiety, a teacher can. . . |
| Handout 2 | How do you *develop* a good relationship with your students? | What can be done to *diminish* issues related to language anxiety? |
| Handout 3 | In order to *develop* a good relationship with his students, he. . . | In order to *diminish* issues related to language anxiety, he. . . |

The sequencing of pre-task activities and tasks as well as the corresponding handouts are shown in Figure 1. In addition to the handouts, an online background survey was used to collect demographic information and an exit survey to collect information about participants' perception of the tasks.

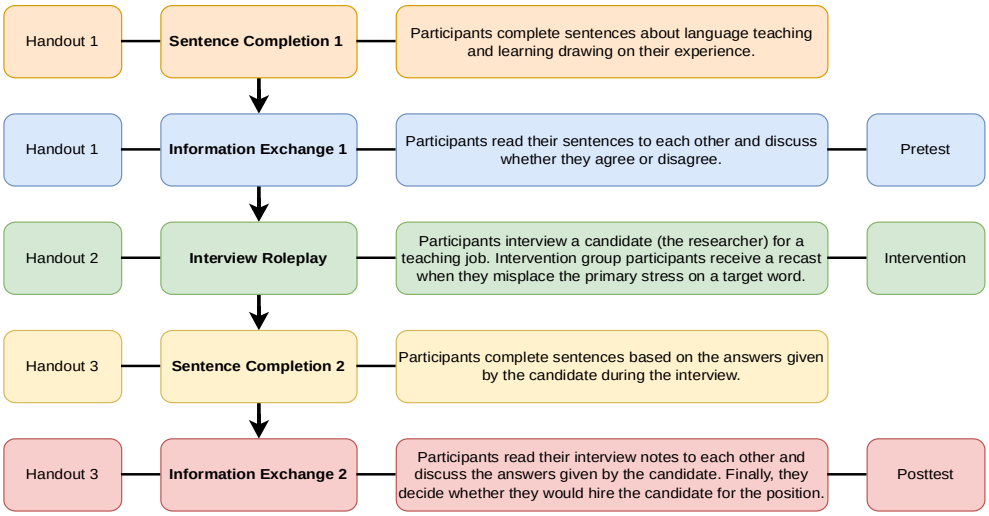


Figure 1 Overview of pre-task activities and tasks

## 3.3. Procedure

Data collection took place in a quiet room at the university where participants studied. Participants were randomly assigned to the control and intervention groups, and they attended data collection sessions in dyads which were also randomly

assigned. After signing the consent form and completing the background survey on a laptop computer, participants were given Handout 1 which required them to complete open-ended sentences about language teaching and learning. The purpose of the sentence-completion activity was to get participants to brainstorm about ideal language teaching practices. Participants were encouraged to fill out the handout by drawing on their own experiences and they were informed that there were no correct or incorrect answers. Student A and Student B in each dyad completed their handout individually. When participants asked about the meaning of a target word, which happened only a few times, the researcher explained the meaning without modeling the pronunciation of the target word. Next, dyads carried out the first information-exchange task which required them to have a brief discussion on the ideas that they generated during the sentence-completion activity. Student A and Student B took turns to read aloud their sentences one at a time and discussed them with each other. The purpose of this task was to encourage participants to share and discuss their ideas on language learning. As the dyads had comparable proficiency in English, the amount of oral production was fairly balanced with no participant dominating the interaction. The dyads' production during the first information-exchange task formed the pretest data. Next, the dyads carried out the interview task with the researcher. They assumed the role of interviewers and interviewed the researcher, who pretended to be an applicant for a teaching position. During the interview task, the dyads took turns to ask the researcher the questions listed on Handout 2. Participants in the intervention group received a recast when they produced a target word with misplaced primary stress with an average of around six recasts per participant. This rate of recast provision was stable across all sessions. As participants joined the sessions in dyads, they were also exposed to the recasts that their partner received. Participants in the control group carried out the same interview task; however, they did not receive corrective feedback. The recasts used in the current study were declarative and single-word recasts (Sheen, 2006). They were initially provided in isolation and then incorporated into the response that immediately followed. Following earlier recommendations by Goo (2012), participants were not given the opportunity for self-correction. The purpose of controlling modified output was to have a tighter control over the experiment, as well as to direct participants' attentional resources to recasts rather than their own or their partner's self-correction. Later, when analyzing the data, it was confirmed that there were no instances of modified output under participants' breath. The exchange shown in Example 2 provides an example of the recast provision procedure from the data (capital letters indicate stress placement).

Example 2 Recast provision

Participant: How do you deveLOP good relationships with your students?
Researcher: Ummm. . . deVElop
Participant: uh-huh
Researcher: To deVElop a good relationship, well, I make myself available to them.
Participant: That's good.

After the interview task, participants used Handout 3 and carried out the second sentence-completion activity using their notes from the interview. Finally, they read aloud their interview notes to each other, discussed the answers given by the candidate, and decided whether they would hire the candidate. The interaction that took place during the second information-exchange task formed the posttest data. When carrying out the information-exchange and interview tasks, participants wore a Shure WH20 XLR brand head-worn unidirectional microphone connected to an audio interface. In order to ensure consistency of acoustic measures, the microphone was positioned at 30 degrees off-axis and about 3 cm away from participants' mouth. Audio files were saved as 44,100Hz .wav files. Participants produced the target words only when they read aloud their sentences; in other words, there was no spontaneous production of the instructional targets during the discussion phase. They did not provide pronunciation feedback to each other during the information-exchange tasks, and there were only six instances where participants repeated a target word back-to-back to fix a false start. In these cases, the more target-like production based on the researcher's auditory judgment was selected for acoustic analysis. After completing the language tasks, participants filled out the exit survey. Each data collection session lasted approximately one hour.

## 4. Analysis

### 4.1. Acoustic analyses

Individual mono tracks were loaded onto Praat (Boersma & Weenink, 2018) for coding. The syllable boundaries of target words were marked using auditory information and visual cues from the waveform and spectrogram, following the maximal onset principle (Deterding, 2001). The initial boundary was marked at the onset of the stop release for syllables starting with a stop, at the onset of voicing approximants and nasals, and at the onset of frication for fricatives. The final boundary was marked at the offset of phonation for syllables that ended with a nasal or approximant, at the offset of frication for fricatives, and at the

offset of the release burst for stops. Word-medial stop gaps were treated as part of the stop consonant that immediately followed (see Figure 2 for an example).

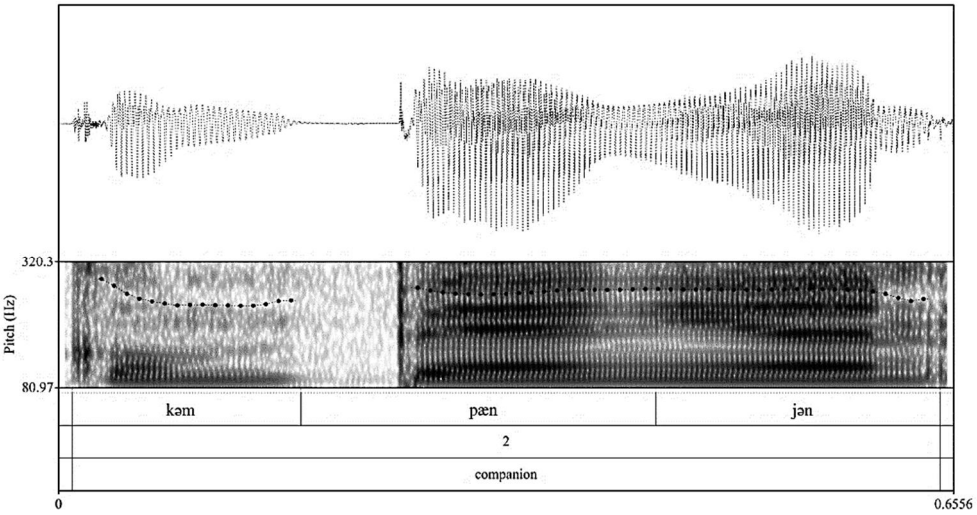


Figure 2 Syllable boundaries for the word companion

Unintelligible words which made it impossible to mark syllable boundaries were removed from the data set along with their corresponding pretest or posttest production. Also, if a participant skipped reading a carrier sentence by mistake resulting in a target word being produced only during the pretest and not on the posttest, or vice versa, that target word was not included in the analysis. Productions with a single phoneme change (e.g., [bəzɪʃən] instead of [pəzɪʃən]) or one missing phoneme (e.g., [kənsɪstən] instead of [kənsɪstənt]) were kept in the data set. Altogether 115 words were removed, and statistical analyses were carried out based on the remaining 1,130 tokens constituting 565 pairs of words produced on the pretest and posttest. The control group produced 261 pairs out of 565 pairs and the intervention group produced 304. The data also showed that 199 target words produced by the intervention group received a recast.

The acoustic analyses were based on syllable duration, intensity, and F0 values to capture the multidimensional nature of lexical stress (Beckman, 1986; Cutler, 2005). A Praat script was used to extract duration, mean intensity in decibels, and mean F0 in hertz from each syllable. F0 measurements represented in linear hertz values were converted to logarithmic semitone values using the f2st function in R (Quené, 2015). As semitones indicate pitch, in other words how F0 is perceived by the human ear, the converted semitone values are referred to as pitch throughout the rest of the paper.

Because suprasegmental features are contrastive and the acoustic properties of what is perceived as a stressed syllable are relative to the acoustic properties of other syllables in the same word (Ladefoged, 2006), a ratio measure was calculated to represent the relative acoustic change in the penultimate syllable. Using such a ratio measure is common practice when analyzing prosodic prominence (e.g., Colantoni et al., 2014; Zuraiq & Sereno, 2021). A ratio measure also allows for a comparison between different acoustic correlates which are measured on different scales such as milliseconds, decibels, and semitones. The ratio measure for each acoustic correlate was calculated by diving the value of the second syllable by the sum of values of the first and third syllables as shown in the following formula: $S_{norm} = S2/(S1+S3)$.

For statistical analyses, linear mixed-effects models were fitted using the lme4 package version 1.1-14 (Bates et al., 2015) working on R version 4.2.0 (R Core Team, 2022). The models included learner-by-time and word-by-time random slopes, as random slopes help with handling by-learner and by-word heteroskedasticity and prevent overconfident results (Barr et al., 2013; Cunnings & Finlayson, 2015). Although linear mixed-effects models are robust against violation of normality (Winter, 2013), normality of duration, intensity, and pitch data were checked visually using q-q plots. All data were normally distributed. The distribution of model residuals for each acoustic correlate was also checked after fitting the models (West et al., 2015). Once again, the q-q plots showed that all model residuals were normally distributed. Finally, the assumptions for linearity and homoscedasticity were checked using residual plots, and it was found that neither homoscedasticity nor linearity were violated.

The models were fitted with *time* and *condition* as fixed effects, and *participants* and *words* as random effects. For each fitted model, the goodness-of-fit measure was calculated as $R^2$ using the piecewiseSEM package version 2.1.2 (Lefcheck, 2016). The estimated marginal means package version 1.8.4-1 (Lenth, 2022) with the Tukey correction was used to obtain pair-wise comparisons.

### 4.2. Listener judgments

Following Saito and Plonsky's (2019) recommendation, the researcher and another expert listener who was a native speaker of American English with a PhD in applied linguistics carried out listener judgments through a two-alternative forced choice (2AFC) experiment created on PsychoPy version 3.0.7 (Peirce et al., 2019). A subset of 60 pretest-posttest word pairs, which amount to 10.62% of the data, were used to prevent listener fatigue. Half of the word pairs were randomly selected from the intervention group's productions and the other half from the control group's productions. Listeners were presented with the orthographic

representation of each word and two recorded versions were played back-to-back with a one-second gap between them. The listeners were then asked to decide whether version A or B had better primary stress placement. Listeners were allowed to replay the two audio files as many times as they wanted before making a decision. Target words were presented in a randomized order to avoid blocking of the data by group or test. Prior to carrying out the experiment, listeners completed a trial experiment with six words which were excluded from the actual experiment. The listening experiment took approximately 25 minutes to complete. Kappa analysis yielded a substantial agreement between the two expert listeners ($\kappa = .76$, $p < .001$).

## 5. Results

### 5.1. Acoustic analyses

The first research question asked how recast provision targeting learners' primary stress errors impact acoustic realization of primary stress. The results for duration, intensity, and pitch are presented in the following sections.

#### 5.1.1. Duration

Descriptive statistics showed that the intervention group produced the second syllables with 5% longer mean duration on the posttest, whereas there was only 0.1% change in control group's performance (see Table 3). The linear mixed-effects model fitted for the analysis of duration yielded a main effect for time and group with an effect size explaining 55% of the variance (see Table 4). The analysis of pairwise contrasts obtained using the emmeans package showed that second syllables produced by the intervention group were statistically longer on the posttest, whereas there was no statistical difference between the pretest and posttest productions by the control group (see Table 9).

Table 3 Descriptive statistics for duration

| Group | Time | *M* | *SD* | *Minimum* | *Maximum* |
|---|---|---|---|---|---|
| Control | Pretest | .636 | .216 | 0.164 | 1.740 |
| | Posttest | .635 | .209 | 0.201 | 1.400 |
| Intervention | Pretest | .598 | .196 | 0.210 | 1.400 |
| | Posttest | .649 | .223 | 0.278 | 1.570 |

Table 4 Linear mixed-effects model estimates for duration

| | *β* estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 0.588 | 0.033 | 22.587 | 17.416 | .000 |
| ConditionControl | 0.037 | 0.016 | 71.315 | 2.206 | .030 |
| TimePosttest | 0.050 | 0.012 | 57.744 | 4.153 | .000 |
| ConditionControl: TimePosttest | -0.053 | 0.017 | 955.927 | -3.106 | .001 |

*Note.* Conditional $R^2$ = .55

Table 5 Descriptive statistics for intensity

| Group | Time | *M* | *SD* | *Minimum* | *Maximum* |
|---|---|---|---|---|---|
| Control | Pretest | .494 | .037 | 0.401 | 0.593 |
| | Posttest | .494 | .039 | 0.389 | 0.617 |
| Intervention | Pretest | .491 | .043 | 0.341 | 0.627 |
| | Posttest | .493 | .039 | 0.395 | 0.612 |

Table 6 Linear mixed-effects model estimates for intensity

| | *β* estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 0.490 | 0.006 | 22.089 | 77.914 | .000 |
| ConditionControl | 0.004 | 0.003 | 81.123 | 1.414 | .161 |
| TimePosttest | 0.002 | 0.002 | 356.520 | 0.948 | .344 |
| ConditionControl: TimePosttest | –0.002 | 0.003 | 810.738 | –0.634 | .527 |

*Note.* Conditional $R^2$ = 0.45

### 5.1.2. Intensity

According to descriptive statistics, there was no difference between the pretest and posttest productions of intensity by the intervention or the control groups (see Table 5). The results of the linear mixed-effects analysis followed the same pattern (see Table 6). Based on pairwise comparisons, neither the intervention group nor the control group produced second syllables with higher intensity on the posttest (see Table 9).

### 5.1.3. Pitch

Descriptive statistics indicated 0.3% gain in pitch for the control group and 0.8% for the intervention group on the posttest (see Table 7). The linear mixed-effects model fitted for the analysis of pitch had a main effect for time and explained 41% of the variance (see Table 8). Pairwise comparisons indicated a statistical difference between the pretest and posttest pitch ratios produced by the intervention group and non-statistical difference between the pretest and posttest ratios produced by the control group. Table 9 provides an overview of pairwise

contrasts for all acoustic correlates. Additionally, the distribution of fitted data extracted from the models can be seen in boxplot form in Figure 3.

Table 7 Descriptive statistics for pitch

| Group | Time | *M* | *SD* | *Minimum* | *Maximum* |
|---|---|---|---|---|---|
| Control | Pretest | 0.512 | 0.036 | 0.377 | 0.677 |
| | Posttest | 0.515 | 0.037 | 0.417 | 0.749 |
| Intervention | Pretest | 0.510 | 0.032 | 0.427 | 0.663 |
| | Posttest | 0.518 | 0.036 | 0.452 | 0.677 |

Table 8 Linear mixed-effects model estimates for pitch

| | *β* estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 0.509 | 0.004 | 45.807 | 122.455 | .000 |
| ConditionControl | 0.001 | 0.004 | 68.657 | 0.405 | .686 |
| TimePosttest | 0.007 | 0.002 | 159.005 | 3.251 | .001 |
| ConditionControl: TimePosttest | -0.005 | 0.003 | 174.845 | -1.551 | .122 |

*Note.* Conditional $R^2$ = 0.41

Table 9 Pairwise pretest-posttest contrasts by experimental condition

| Group | | *β* estimate | *SE* | *df* | *t*-value | *p*-value |
|---|---|---|---|---|---|---|
| Control | Duration | -0.002 | 0.013 | 37.7 | -0.195 | .846 |
| | Intensity | 0.000 | 0.002 | 419 | 0.028 | .977 |
| | Pitch | 0.002 | 0.002 | 43.3 | 0.911 | .367 |
| Intervention | Duration | 0.050 | 0.012 | 32.6 | 4.142 | .000 |
| | Intensity | 0.002 | 0.002 | 357 | 0.948 | .343 |
| | Pitch | 0.007 | 0.002 | 38.4 | 3.243 | .002 |

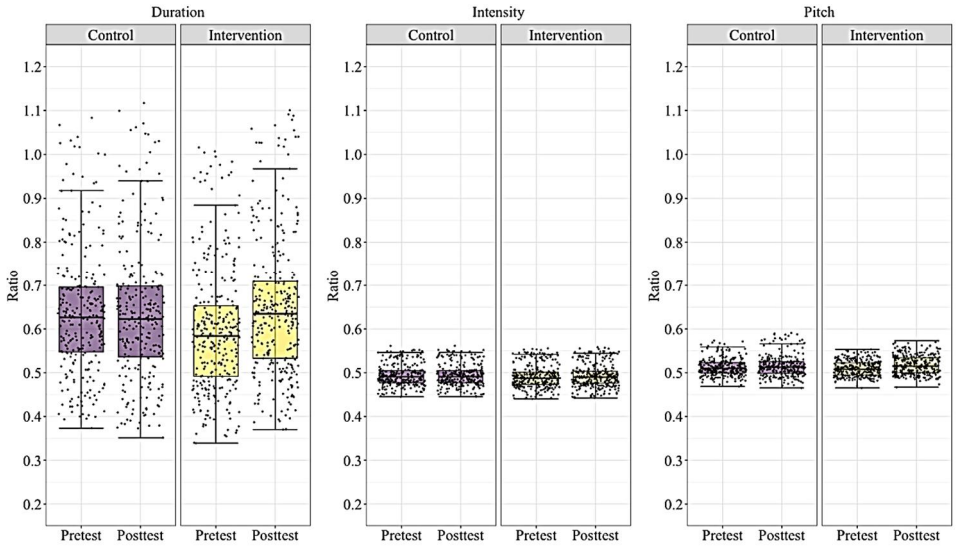


Figure 3 Distribution of fitted data by condition

### 5.1.4. Recast vs. no recast

A subsequent set of analyses was run to investigate the potential acoustic differences between the words that received a recast and words that did not as produced by the intervention group. The results showed that intervention group participants produced the second syllables of the words that received a recast with 7.5% longer duration on the posttest, which was statistically significant, whereas their productions of the words that did not receive a recast were not statistically different on the posttest. There was no statistical effect for the intensity ratios regardless of whether the words received a recast or not. Finally, the intervention group produced the second syllables of the words that received a recast with 1% statistically higher pitch on the posttest, while there was no statistical difference for the words that did not receive a recast. Table 10 provides an overview of the pairwise comparisons for the words that received a recast and the words that did not.

Table 10 Pairwise pretest-posttest contrasts by feedback provision

| Group | | *β* estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No recast | Duration | -0.001 | 0.021 | 62.6 | -0.049 | .960 |
| | Intensity | 0.003 | 0.004 | 66.7 | 0.739 | .462 |
| | Pitch | 0.002 | 0.003 | 72.3 | 0.746 | .458 |
| Recast | Duration | 0.075 | 0.016 | 26.4 | 4.639 | .000 |
| | Intensity | 0.001 | 0.003 | 27.3 | 0.509 | .615 |
| | Pitch | 0.010 | 0.002 | 30.9 | 3.558 | .001 |

## 5.2. Listener judgments

The second research question asked how changes in acoustic measures impact expert listener judgments. Expert listener judgments showed that the intervention group produced 77% of the words with relatively better stress placement on the posttest. On the other hand, the control group data showed that 57% of the words had a relatively better stress placement on the posttest. There was no agreement between the two raters for 13% of the words produced by the intervention group and 10% of the words produced by the control group (see Figure 4). The results of a two-tailed binomial test run on R indicated that the intervention group's 23-77 ratio favoring posttest productions was statistically significant ($p$ = .005, *95% CI* = [-0.577, -0.900]), whereas the control group's 43-57 ratio was not ($p$ = .585, *95% CI* = [-0.374, -0.745]).

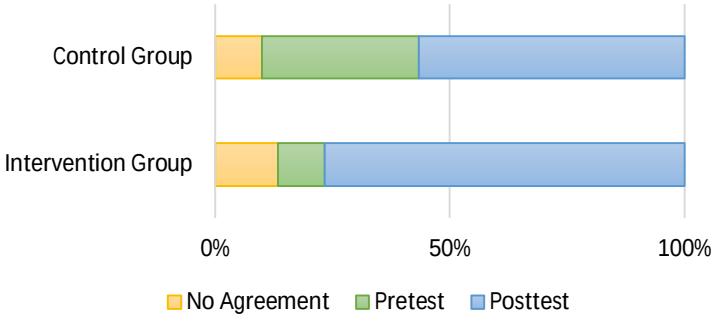


Figure 4 Percentage of words produced with better primary stress based on listener judgments

## 6. Discussion

Motivated by recent interactionist L2 phonology studies, the goal of the current study was to investigate the effects of recasts on the production of primary stress. The first research question focused on the relationship between recasts and the potential improvements in acoustic realization of primary stress. The results indicated that recasts led to a significant increase in syllable duration on the posttest. There was also a significant increase in pitch values; however, the distribution of fitted data showed that the gains were much less compared to the gains in duration. Subsequent analyses focusing on the productions of the intervention group showed that the words that received a recast were produced with statistically longer duration and higher pitch, while there were no statistical changes in the production of the words that did not receive a recast. In other words, the intervention group modified their production of primary stress only for the target words that were brought to focus by recasts. The second research question focused on the relationship between recasts and target-like production based on listener judgments. The results showed that the positive effect of recasts on acoustic realization of primary stress was supported by expert listener judgments. Considering the role of duration as a robust correlate of lexical stress (Gordon & Roettger, 2017; Kochanski et al., 2005) as well as L1 Arabic speakers' tendency to utilize duration for realization of stress in Arabic and English (de Jong & Zawaydeh, 2002; Zuraiq & Sereno, 2021), the findings indicate a positive relationship between recasts and primary stress development. In this way, the findings also lend support to the recent interactionist studies which have argued for the positive role of recasts in facilitating productive development of L2 phonology (Bryfonski & Ma, 2020; Saito, 2013, 2015).

Another important aspect of the current study is that it is one of the few studies documenting the acoustic changes that corrective feedback triggers in the realization of lexical stress. The current study extends the findings by Parlak and Ziegler (2017), which showed that upon receiving recasts, L1 Arabic speakers increased syllable duration to improve their primary stress production. Although their results were not statistically significant at the group level, subsequent analyses focusing on a subset of three-syllable words produced significant gains for duration. Taken together, the findings of Parlak and Ziegler (2017) and the current study suggest that recasts targeting primary stress errors push L1 Arabic speakers to utilize duration to improve their primary stress placement in English. One possible reason for the larger impact of recasts on syllable duration compared to intensity and pitch could be the cross-linguistic influence on the perception and production of stress in a second language, a factor that has been highlighted by earlier studies (e.g., Archibald, 1997; Nguyễn et al., 2008; Zhang et al., 2008). As vowel duration plays a key role in the production of stress in Arabic (Alrajeh, 2011; de Jong & Zawaydeh, 2002), it is possible that participants exclusively focused on the durational differences between the recasts that they received and their own production, while ignoring the differences in intensity and pitch. Alternatively, they may have perceived the prominence of the stressed syllable as a product of two or all three acoustics correlates; and yet, they may have utilized duration to bridge the gap between their production and the recast. This particular focus on duration could also be the reason for the distributional differences in the fitted data for the three acoustic correlates, which showed that the variation in participants' production of duration was larger than their production of intensity and pitch (see Figure 3). Previous research has shown that L1 Arabic speakers have a tendency to focus on durational elements when producing vowels in English, including the contrast between tense and lax vowels (Munro, 1993), and that there is cross-linguistic influence from Arabic affecting the phonetic quality of the vowels produced in L2 English (Flege & Port, 1981). Similarly, participants in the current study may have exclusively focused on durational effects to improve their production of primary stress, leading to higher levels of fluctuation and a wider range of distribution of duration values.

Although the current study did not compare different types of corrective feedback, the interactions between the researcher and participants during the interview task highlight the suitability of recasts when addressing pronunciation issues. Recasts provided by the researcher did not break the natural flow of interaction or cause participants to stop and ask metalinguistic questions. As discussed earlier, recasts unobtrusively provided participants with a model that they could juxtapose with their production without pushing them for repetition or self-correction. In this way, the recasts in the current study were meaning-

focused and non-threatening. However, it is also necessary to highlight that the procedure for recast provision was intensive and relatively explicit. Recasts were produced with extra prosodic emphasis, which undoubtedly increased the perceived prominence of the target syllable. Also, the double provision method, which presented recasts first in isolation and then as part of the response given to participants, most likely made the recasts highly salient (Philp, 2003) and increased their noticeability. The current study did not adopt a direct measure of noticing; however, it utilized a retrospective questionnaire to gain insights into participants' perceptions of the interactions. One of the survey questions was "Did the researcher correct your English during the interview task? If yes, what did he correct and how did you react? Please explain." The number of participants who responded with a "Yes" was 74% for the intervention group and only 3% for the control group. Responses to the open-ended part of the question by the intervention group included comments about receiving feedback on their pronunciation (e.g., "yes, he corrected my pronunciation by repeating the same word again"). Although the questionnaire did not include a question about the noticing of recasts provided to peers, six participants from the intervention group indicated that their partner also received corrective feedback (e.g., "he also corrected my partner's pronunciation"). The comments suggest that participants noticed peer-directed recasts and possibly benefited from it as well. Participants also mentioned specific target words as part of their response (e.g., "yes, he corrected for me the incorrect word like perception;" "yes, when I say consensus wrong, he told me the right way to say it and I say it correctly;" "yes, the word contention;" "dynamic"). These responses can be taken as evidence that at least some level of noticing of recasts occurred, which is an important factor affecting the usefulness of corrective feedback (Ellis, 2016). Furthermore, the responses are aligned with earlier research which argued that learners more readily notice corrective feedback when it targets pronunciation errors (Carpenter et al., 2006; Mackey et al., 2000).

It is also worth mentioning that the effectiveness of corrective feedback depends on a number of factors, one of which is the target language feature (Kartchava & Ammar, 2014; Mackey et al., 2000; Saito & Lyster, 2012). Lexical stress is relatively easier to produce compared to other speech phenomena such as aspiration, consonant clustering, or novel segmental features. Stress can be realized through producing a syllable with longer duration and higher vocal energy, which are relatively easier to accomplish for many L2 speakers compared to producing novel segmental features with complex articulation (e.g., consonant clusters in Russian or clicks in Xhosa). Therefore, it is possible that the relative ease of production of lexical stress contributed to the positive results. Future studies can shed light on how implicit corrective feedback works on more marked phonological features that require complex articulatory movements.

## 7. Limitations and future research

As with most research studies, a few limitations should be taken into account when interpreting the results. For one thing, the current study did not explore whether recasts led to perceptual gains. Understanding the nature of the relationship between corrective feedback and perception of phonological features is necessary as earlier research highlighted the role of perception as a precondition to production (Best & Tyler, 2007; Flege, 1995; Leung et al., 2021). There are already some interaction studies which have shown the positive impact of recasts on the perception of segmental features (Lee & Lyster, 2016). To determine whether similar effects are present in the case of lexical stress or other suprasegmental features, future research needs to examine the potential changes that may occur in both perception and production.

Another limitation of the current study is the length of the intervention. The data collection phase was limited to one hour. Although L2 learners may be able to improve an already developing aspect of their L2 with the help of corrective feedback provided during a brief period of interaction, longer periods of sustained exposure to the target language as well as feedback would create the ideal condition for meaningful changes to occur. This is particularly the case for pronunciation development, which usually lags behind other aspects of language development despite years of active language use (Derwing & Munro, 2015). Also, it was not possible to measure sustained effects of recasts as the current study did not have a delayed posttest due to logistical reasons. The data were collected right before the summer break and participants indicated that they would not be available for a delayed posttest. There is evidence that, in some cases, the positive impact of recasts may be delayed (see Mackey & Goo, 2007). As such, it is possible that facilitative impact of recasts on production of stress observed on the immediate posttest could increase or at least be sustained after some time. That said, without delayed posttest data, it is not possible to make firm conclusions about the delayed effects. In light of all this, the current study should be considered as a preliminary one with the findings supporting the facilitative role of recasts based on immediate production.

Finally, the current study did not measure the impact of recasts on participants' production of untrained vocabulary, meaning new words that were not part of the intervention. Saito (2015) suggests that positive changes in both trained and untrained vocabulary would be stronger evidence for phonological development. The argument is based on the premise that inability to transfer phonological development to untrained vocabulary indicates lexical learning, which would be limited to the set of vocabulary used in that particular study. This suggestion is definitely meaningful in the case of segmental features. However, stress placement

in English can in fact be lexical, which was also the case for the target words in the current study. As mentioned earlier, a sizable number of lexical stress patterns in English need to be learned on a case-by-case basis. Yet, stress in English can also be paradigmatic and follow a transferrable pattern. Therefore, future research is necessary to understand developmental patterns by examining whether learners can transfer gains to untrained vocabulary with the same stress pattern.

## 8. Conclusion

The current study is among the few studies that explored the impact of corrective feedback on the development of a suprasegmental feature through acoustic analysis and expert listener judgments. The findings indicated that recasts provided during a communicative activity had a positive impact on the immediate production of primary stress in the form of longer syllable duration, which also had a positive impact on expert listener judgments. The findings lend further support to the growing number of interaction studies arguing for the usefulness of recasts when addressing pronunciation issues and encourage the adoption of the interactionist approach to investigate different aspects of L2 phonological development.

## Acknowledgements

The data reported in this paper mainly come from my PhD dissertation completed at Lancaster University, United Kingdom, with the exception of the listener judgments which were carried out subsequently. I would like to thank my thesis supervisors Sam Kirkham, Alison Mackey, and Jenefer Philp for their guidance on this research project. I would also like to thank my dissertation committee, Adrian Leemann and Andrea Révész, for their comments and suggestions. Last but not least, many thanks to the anonymous SSLLT reviewers and the editorial team for their constructive feedback. Any remaining errors are my own.

## References


Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014). Acquiring the phonetics and phonology of English word stress: Comparing learners from different L1 backgrounds. *Concordia Working Papers in Applied Linguistics*, *5*, 19-35.

Alrajeh, N. (2011). Digital processing of stress in standard Arabic. *Journal of King Saud University-Languages and Translation*, *23*(2), 65-68. https://doi.org/10.1016/j.jksult.2011.03.001

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation. *Linguistics, 35*(1), 167-181. https://doi.org/10.1515/ling.1997.35.1.167

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.48550/arXiv. 1406.5823

Beckman, M. (1986). *Stress and non-stress accent.* Foris. https://doi.org/10.1515/9783110874020

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-24). John Benjamins.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.28) [Computer program]. http://www.praat.org/

Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning, 52*(1), 1-42. https://doi.org/10.1111/1467-9922.00176

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*(4), 436-458. https://doi.org/10.1177/1362168814563200

Bryfonski, L., & Ma, X. (2020). Effects of implicit versus explicit corrective feedback on Mandarin tone acquisition in a SCMC learning environment. *Studies in Second Language Acquisition, 42*, 61-88. https://doi.org/10.1017/S0272263119000317

Carpenter, H., Jeon, K. S., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition, 28,* 209-236. https://doi.org/10.1017/ S0272263106060104

Colantoni, L., Marasco, O., Steele, J., & Sunara, S. (2014). Learning to realize prosodic prominence in L2 French and Spanish. In R. T Miller et al. (Eds.), *Selected proceedings of the 2012 second language research forum: Building bridges between disciplines* (pp. 15-29). Cascadilla Proceedings Project.

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159-181). Routledge.

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264-289). Blackwell.

de Jong, K., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30,* 53-75. https://doi.org/10.1006/jpho.2001.0151

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research.* John Benjamins.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics, 29,* 217-230.

Educational Testing Service. (2010). Linking TOEFL iBT scores to IELTS scores: A research report. Available at https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research*, *20*(3), 405-428. https://doi.org/10.1177/1362168816628627

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423. https://doi.org/10.2307/3588487

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425-442. https://doi.org/10.1017/S0142716400066029

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, *24*(2), 125-146. https://doi.org/10.1177/002383098102400202

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, *9*(2), 198-217. https://doi.org/10.1093/applin/9.2.198

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, *19*(1), 3-17. https://doi.org/10.1075/aila.19.03gas

Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 180-206). Routledge.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*, 445-474. https://doi.org/10.1017/S0272263112000149

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, *35*(1), 127-165. https://doi.org/10.1017/S0272263112000708

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1), 1-21. https://doi.org/10.1515/lingvan-2017-0007

Hellmuth, S. (2013). Phonology. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 45-70). Oxford University Press.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, *34*(3), 475-505. https://doi.org/10.1017/S0272263112000150

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*(4), 554-566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kartchava, E. & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research, 18*, 428-452. https://doi.org/10.1177/1362168813519373

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038-1054. https://doi.org/10.1121/1.1923349

Ladefoged, P. (2006). *A course in phonetics*. Thomson Wadsworth.

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*(1), 35-64. https://doi.org/10.1017/S0272263115000194

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25*(1), 37-63. https://doi.org/10.1017/S0272263103000020

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573-579. https://doi.org/10.1111/2041-210X.12512

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means.* Retrieved from https://CRAN.R-project.org/package=emmeans

Leung, A. H. C., Young-Scholten, M., Almurashi, W., Ghadanfari, S., Nash, C., & Outhwaite, O. (2021). (Mis) perception of consonant clusters and short vowels in English as a foreign language. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2021-0030

Lewis, C., & Deterding, D. (2018). Word stress and pronunciation teaching in English as a lingua franca contexts. *CATESOL Journal*, *30*(1), 161-176.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, *24*(4), 649-666. https://doi.org/10.2307/3587113

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Long, M. H. (2007). *Problems in SLA.* Lawrence Erlbaum.

Mackey, A., Abbuhl, R., & Gass, S. M. (2012). Interactionist approach. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 7-23). Routledge.

Mackey, A., Gass, S. M., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*(4), 471-497. https://doi.org/10.1017/S0272263100004010

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford University Press.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39-66. https://doi.org/10.1177/002383099303600103

Mustafawi, E. (2018). Arabic phonology. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 11-31). Routledge.

Nguyễn, T. A. T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158-190. https://doi.org/10.1016/j.wocn.2007.09.001

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition, 39*(2), 257-285. https://doi.org/10.1017/S0272263116000310

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, & E., Lindeløv, J. (2019). *PsychoPy2: Experiments in behavior made easy*. Behavior Research Methods. https://doi.org/10.3758/s13428-018-01193-y

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25,* 99-126. https://doi.org/10.1017/S0272263103000044

Quené, H. (2015). *Package 'hqmisc'*. https://cran.r-project.org/web/packages/hqmisc/hqmisc.pdf

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.0). R Foundation for Statistical Computing. https://www.R-project.org/

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*(1), 1-29. https://doi.org/10.1017/S0272263112000666

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics*, *36*(2), 377-409. https://doi.org/10.1017/S0142716413000271

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*, 595-633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652-708. https://doi.org/10.1111/lang.12345

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217-240. https://doi.org/10.1017/S0142716414000502

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition*, *36*(4), 647-680. https://doi.org/10.1017/S0272263114000114

Scheuer, S., & Horgues, C. (2019). Corrective feedback in English/French spoken tandem interactions. In C. Tardieu & C. Horgues (Eds.), *Redefining tandem language and culture learning in higher education* (pp. 147-160). Routledge.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1-32). Cambridge University Press.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*(3), 263-300. https://doi.org/10.1191/1362168804lr146oa

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*(4), 361-392. https://doi.org/10.1191/1362168806lr203oa

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-483). Lawrence Erlbaum.

Watson, J. C. (2007). *The phonology and morphology of Arabic.* Oxford University Press.

West, B. T., Welch, K. B., & Galecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software*. CRC Press.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. Retrieved from https://doi.org/10.48550/arXiv.1308.5499

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498-4513. https://doi.org/10.1121/1.2902165

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, *38*(3), 553-586. https://doi.org/10.1017/S027226311500025X

Zuraiq, W., & Sereno, J. (2021). Production of English lexical stress by Arabic speakers. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 290-311). Cambridge University Press.