JASA ARTICLE

# Dutch listeners' perception of English lexical stress: A cue-weighting approach

Annie Tremblay,[1,a)] Mirjam Broersma,[2] Yuyu Zeng,[1] Hyoju Kim,[1] Jinmyung Lee,[1] and Seulgi Shin[1]
[1]Department of Linguistics, University of Kansas, Lawrence, Kansas 66045, USA
[2]Centre for Language Studies, Radboud University, 6500 HD Nijmegen, The Netherlands

**ABSTRACT:**
We investigate whether acoustic cue weightings are transferred from the native language to the second language [research question 1 (RQ1)], how cue weightings change with increasing second-language proficiency (RQ2), and whether individual cues are used independently or together in the second language (RQ3). Vowel reduction is a strong cue to lexical stress in English but not Dutch. Native English listeners and Dutch second-language learners of English completed a cue-weighting stress perception experiment. Participants heard sentence-final pitch-accented auditory stimuli and identified them as *DEsert* (initial stress) or *deSSERT* (final stress). The stimuli were manipulated in seven steps from initial to final stress, manipulating two dimensions at a time: vowel quality and pitch, vowel quality and duration, and pitch and duration (other dimensions neutralized). Dutch listeners relied less on vowel quality and more on pitch than English listeners, with Dutch listeners' sensitivity to vowel quality increasing with English proficiency but their sensitivity to pitch not varying with proficiency; Dutch listeners evidenced similar or weaker reliance on duration than did English listeners, and their sensitivity to duration increased with proficiency; and Dutch listeners' use of pitch and duration were positively related. These results provide general support for a cue-based transfer approach to the perception of lexical stress. *© 2021 Acoustical Society of America*.
https://doi.org/10.1121/10.0005086

(Received 11 November 2020; revised 5 April 2021; accepted 8 May 2021; published online 2 June 2021)

[Editor: Benjamin V. Tucker] Pages: 3703–3714

## I. INTRODUCTION

It is well established that the sound system of the native (i.e., first) language (L1) influences how listeners perceive speech in a second or foreign language (L2). In the domain of prosody, however, the precise nature of this influence has yet to be understood and adequately explained. One influential approach that has sought to understand L1 effects on the perception of segmental and prosodic contrasts is the cue-weighting theory of speech perception (Francis *et al.*, 2000; Francis and Nusbaum, 2002; Holt and Lotto, 2006). This theory stipulates that the perceptual learning of speech in the L1 and in the L2 is served by statistical computation and selective attention mechanisms: Listeners extract the regularity of acoustic cues to linguistically meaningful contrasts as a function of the distribution of these cues in the signal and direct their attention to those cues that are the most informative for identifying linguistic contrasts. This theory proposes that listeners use a variety of acoustic cues simultaneously to perceive sound contrasts but weight these cues as a function of their informativeness for signaling lexical contrasts. Because cues are weighted differently across languages, L1 effects on the perception of L2 sound contrasts are attributed to listeners' transfer of their cue weightings from the L1 to the L2, also referred to as the *cue-weighting transfer hypothesis* [for examples of L2 studies within such a framework, see Chrabaszcz *et al.* (2014), Ingvalson *et al.* (2012), Iverson *et al.* (2003), and Zhang and Francis (2010)].

The present study provides a direct test of the cue-weighting transfer hypothesis in the domain of prosody by newly addressing the following three research questions:

(RQ1) How does L2 learners' knowledge of acoustic cues to lexical stress in the L1 modulate their perception of lexical stress in the L2?
(RQ2) How does L2 learners' reliance on acoustic cues to lexical stress in the L2 develop in relation to the functional weight of these cues in the L1 and in the L2?
(RQ3) Is L2 learners' reliance on different acoustic cues to lexical stress independent (i.e., sensitivity to one cue does not predict sensitivity to another cue) or interrelated (i.e., sensitivity to one cue predicts sensitivity to another cue)?

More specifically, this study provides a novel investigation of how Dutch L2 learners of English and native English listeners weight vowel quality, pitch, and duration cues to English lexical stress (RQ1), how Dutch listeners' weighting of acoustic cues to English stress changes with increasing English proficiency (RQ2), and whether Dutch listeners' uses of vowel quality, pitch, and duration cues to English lexical stress are independent or interrelated (RQ3).

Previous research has suggested that Dutch listeners rely more on suprasegmental cues to lexical stress than do English listeners (Cooper *et al.*, 2002; Cutler *et al.*, 2007).

[a)]Electronic mail: atrembla@ku.edu

 © 2021 Acoustical Society of America 

For example, compared to native English listeners, Dutch L2 learners of English are more accurate at selecting the correct word continuation after hearing fragments (e.g., *mus*–) that belong to either a word with primary stress on the initial syllable (e.g., "music") or a word with primary stress on the second syllable [e.g., "museum"; experiment 3 of Cooper *et al*. (2002)], with the fragments not containing segmental cues to stress. These findings suggest that Dutch listeners make greater use of suprasegmental cues to English lexical stress than native English listeners. In a follow-up study, Cutler *et al*. (2007) report significant correlations between these L2 learners' responses and suprasegmental properties of the word fragments, with none of these correlations being significant for the native listeners. These findings indicate that Dutch listeners are more likely to attend to the suprasegmental properties of the fragments when selecting a word continuation. However, these studies have not explicitly tested Dutch listeners' cue weightings, raising the question of exactly how Dutch listeners weight segmental and suprasegmental cues to lexical stress in English (RQ1).

From a learnability perspective, it is also unclear whether L2 learners' reliance on individual acoustic cues develops in parallel or asymmetrically with increasing proficiency depending on the functional weight of these cues in the L1 and the L2 (RQ2). Some L2 speech processing studies in the *segmental* domain have suggested that it is easier to learn to use cues that play an important role in the L2 but not the L1 than it is to suppress the use of cues that play an important role in the L1 but not the L2 (Tremblay and Spinelli, 2014; Weber and Cutler, 2006). However, this finding has not been replicated in a sufficiently large number of studies for firm conclusions on the learnability of acoustic cues to be reached. Importantly, it is also unclear whether this finding extends to the domain of *prosody*, such as the perception of lexical stress.

Last but not least, there is currently a limited understanding of whether L2 learners' weightings of different acoustic cues to lexical stress are independent or interrelated (RQ3). Some studies on the perception of *segmental* contrasts have found that L2 learners' increased reliance on one cue to the contrast tended to co-vary with a decreased reliance on another cue to the same contrast (e.g., Flege *et al*., 1997; Schertz *et al*., 2015). Other studies, however, do not report a systematic relationship between L2 learners' weightings of different cues to the same segmental contrast [e.g., the individual cue weightings reported in Tables B2 and B3 of Kim *et al*. (2018) are negatively related, but only one in eight correlations (that we conducted) reached significance]. Yet other studies have found a positive relationship between listeners' use of different cues to the same segmental contrast (e.g., Clayards, 2018; Hazan and Rosen, 1991; Shultz *et al*., 2012), suggesting that some L2 learners may be better at perceiving phonetic information than others. For prosodic phenomena such as lexical stress, it remains to be seen whether L2 learners' uses of different acoustic cues are independent or interrelated, and if the latter, what is the directionality of the observed relationships.

By remedying these gaps (RQ1–RQ3), the present study will not only provide a robust test of whether listeners transfer their relative reliance on acoustic cues to lexical stress from the L1 to the L2—thereby providing another test of the cue-weighting transfer hypothesis—but also help elucidate the nature of the between-learner variability that characterizes L2 learners' cue weightings. Importantly, answering these questions will also contribute to determining whether L2 learners' perception of lexical stress tends to be unidimensional (strong within-learner reliance on a single cue and weak or no reliance on other cues) or multidimensional (moderate-to-strong within-learner reliance on several cues).

American English (hereafter, English) has lexical stress contrasts (e.g., *DEsert* vs *deSSERT*, where the capitalized letters represent the stressed syllables) (Halle and Vergnaud, 1987; Hammond, 1999). The most important acoustic correlate of lexical stress in English is vowel quality: Stressed syllables contain a full (i.e., unreduced) vowel (e.g., the first syllable of *DEsert* contains /ɛ/), and unstressed syllables tend to contain a reduced vowel (e.g., the first syllable of *deSSERT* contains /ə/), with reduced vowels differing from full vowels in having a more centralized place of articulation and a shorter duration (Gay, 1978; Lindblom, 1963). Another important acoustic correlate of lexical stress in English is duration: Across intonational realizations, stressed syllables are consistently longer than unstressed syllables, with this acoustic correlate co-varying with vowel quality (Beckman and Edwards, 1994) and being attenuated in phrase-final position due to phrase-final lengthening (Nakatani *et al*., 1981; Turk and Shattuck-Hufnagel, 2007). Importantly, however, there is no one-to-one relationship between pitch and lexical stress in English: Words can have different intonational pitch accents (e.g., H*, L + H*, L*, L*+H) and thus different fundamental frequency (F0) patterns depending on discourse structure information, with the realization of these pitch accents depending on the position where words are elicited in relation to phrase and tone boundaries (e.g., pitch descends in declarative utterances); words can also be unaccented (e.g., verbs often do not receive a pitch accent) or deaccented after a word with a contrastive pitch accent (Beckman, 1986; Beckman and Pierrehumbert, 1986; Ladd, 2012; Pierrehumbert, 1980; Pierrehumbert and Hirschberg, 1990). Hence, pitch can only be interpreted as a cue to lexical stress if sentence intonation is taken into consideration (Beier and Ferreira, 2018; Brown *et al*., 2016; Cutler and Fox, 1977; Ortega-Llebaria *et al*., 2019; Shields *et al*., 1974).

Dutch also has lexical stress contrasts (e.g., *CAnon* "canon, round song" vs *kaNON* "cannon") (Gussenhoven, 2008), with stressed syllables being longer than unstressed syllables across intonational realizations (Sluijter and van Heuven, 1996a) but with durational differences being similarly attenuated in phrase-final position (Cambier-Langeveld, 1997). Dutch words can also have different intonational pitch accents (e.g., H*, H*L, H*!H, L*, L*H), with the pitch accent being anchored to the lexically stressed syllable and its realization depending on the position where it is elicited in relation to phrase and tone boundaries, and words can similarly be unaccented (e.g., verbs) or deaccented after a contrastive accent (Gussenhoven, 2004; Krahmer and Swerts, 2001). Thus, it is also the case that pitch cues to lexical stress must be interpreted in the context of a particular intonation in Dutch. Crucially, Dutch differs from English in that unstressed vowels are less reduced in Dutch than in English (Sluijter and van Heuven, 1996a), vowel reduction thus not being as strong of a correlate of lexical stress in Dutch and suprasegmental cues to stress (pitch, duration, intensity) thus having a greater *relative* weight in Dutch than in English. Accordingly, Dutch L2 learners of English produce a lesser degree of vowel reduction in English compared to native English speakers (Braun *et al.*, 2008).

Speech perception studies have shown that vowel quality is the strongest cue to lexical stress in English, with English listeners' stress perception depending more on vowel quality than on pitch, duration, or intensity (Chrabaszcz *et al.*, 2014; Zhang and Francis, 2010).[1] Spoken word recognition studies have also found that English listeners are more sensitive to lexical stress when unstressed vowels are reduced than when they are full (Connell *et al.*, 2018; Fear *et al.*, 1995; Small *et al.*, 1988), with English listeners nonetheless using co-occurring suprasegmental cues to stress in lexical access in the absence of segmental cues (Connell *et al.*, 2018; Cooper *et al.*, 2002; Jesse *et al.*, 2017; Tremblay, 2008). Native Dutch listeners also use stress in lexical access (van Donselaar *et al.*, 2005; Reinisch *et al.*, 2010), but unlike English listeners, Dutch listeners rely more on duration than on vowel quality when perceiving lexical stress in Dutch (van Heuven and de Jonge, 2011), as tested with unaccented words that did not undergo phrase-final lengthening; however, the relative weight of cues to Dutch lexical stress in other intonational contexts (e.g., in sentence-final position and in words with a pitch accent, where pitch cues to lexical stress are prominent) has not, to our knowledge, been assessed. Furthermore, Dutch listeners' identification of Dutch words is more strongly impacted by stress errors realized with co-occurring suprasegmental cues than native English listeners' identification of English words (van Leyden and van Heuven, 1996). Last but not least, as discussed earlier, Dutch L2 learners of English make greater use of co-occurring suprasegmental cues to stress when identifying English words than do native English listeners [experiment 3 of Cooper *et al.* (2002)]. These findings suggest that Dutch listeners transfer their weightings of acoustic cues to lexical stress from Dutch to English. However, this hypothesis has not been explicitly tested.

The present study uses a cue-weighting stress perception experiment to investigate how Dutch L2 learners of English weight acoustic cues to lexical stress in English. Participants heard stimuli with a nuclear pitch accent in sentence-final position, where pitch cues to lexical stress are prominent (because of the nuclear pitch accent) and where duration cues to lexical stress are attenuated (because of phrase-final lengthening). The stimuli were manipulated in seven acoustically equidistant steps from word-initial stress to word-final stress, orthogonally manipulating two dimensions at a time: pitch and vowel quality, duration and vowel quality, and pitch and duration (nonmanipulated dimensions were neutralized). Given previous findings on Dutch listeners' perception of lexical stress (Cooper *et al.*, 2002; van Heuven and de Jonge, 2011; van Leyden and van Heuven, 1996), we predict that Dutch L2 learners of English will rely less on vowel quality cues and more on suprasegmental cues to English lexical stress compared to native English listeners (RQ1). Given the prominence of pitch cues and the attenuation of duration cues in the current stimuli, we expect that Dutch listeners' perception of lexical stress will be more strongly influenced by pitch than by duration (cf. van Heuven and de Jonge, 2011). In light of the research on listeners' transfer of L1 cues and learning of L2 cues (Tremblay and Spinelli, 2014; Weber and Cutler, 2006), we predict that Dutch L2 learners of English will become more sensitive to vowel quality cues to English lexical stress as their proficiency in English increases, but their use of suprasegmental cues to English stress will not be modulated by their proficiency, resulting in a cue weighting where suprasegmental cues continue to play a relatively more important role than for native English listeners (RQ2). Finally, despite the inconsistent findings reported in the literature on the relationships among L2 learners' weightings of acoustic cues to segmental contrasts (e.g., Flege *et al.*, 1997; Schertz *et al.*, 2015), and in view of the research showing that L2 learners as a group can rely on several cues to English lexical stress (e.g., Zhang and Francis, 2010), we anticipate that Dutch listeners' perception of English stress will be multidimensional rather than unidimensional, with L2 learners' strong reliance on one cue not resulting in their weak reliance on another cue (RQ3).

## II. METHOD

### A. Participants

Twenty-seven native English listeners (mean age: 20.8, standard deviation [SD]: 3) and 40 Dutch L2 learners of English (mean age: 21.5, SD: 3) participated in this study. The native English listeners were tested at an American university, and the Dutch L2 learners of English were tested at a Dutch university. All participants completed a detailed language background questionnaire. None of the participants reported having speech or learning impairments. The native English listeners reported having native-speaking English parents, hearing and speaking English in the household before the age of 18, and spending their childhood and teenage years in the United States. The Dutch L2 learners of English reported having native-speaking Dutch parents, hearing and speaking Dutch in the household before the age of 18, and spending their childhood and teenage years in the Netherlands. The L2 learners completed the Lexical Test for Advanced Learners of English (LexTALE) as a proficiency measure (Lemhöfer and Broersma, 2012).[2] Dutch listeners'

TABLE I. L2 learners' experience with and proficiency in English.

| | Age of onset of English Instruction | Years of English instruction | Months in English-speaking country | Use of English (%) | LexTALE score (%) |
|---|---|---|---|---|---|
| Mean (SD) | 11.0 (1.1) | 7.5 (1.7) | 2.3 (5.7) | 20.5 (15.2) | 78.2 (10.3) |
| Range | 9–14 | 4–12 | 0–32 | 4–60 | 64–99 |

experience with English and their English proficiency scores are summarized in Table I. Given the range of LexTALE scores shown in Table I, our L2 learners were at an "upper intermediate" or "upper and lower advanced/proficient user" level of proficiency in English [corresponding levels from the Common European Framework; Lemhöfer and Broersma (2012), Appendix C].

### B. Materials

The auditory stimuli used in the stress perception experiment were produced by a female native speaker of American English (age 26). Four repetitions of *DEsert-deSSERT* were elicited and recorded in the carrier sentence *Click on* ___ (one repetition per block, each with a mixture of words with initial stress and words with final stress). Given this intonational context, the target words were produced with a nuclear pitch accent, where pitch cues to lexical stress are prominent, and they were elicited in sentence-final position, where duration cues to lexical stress are attenuated because of phrase-final lengthening. The speaker was asked to produce a brief pause between the preposition *on* and the target word. The speech productions were audiorecorded in an anechoic chamber using an Electro-Voice (Burnsville, MN) N/D 767 cardioid microphone and a Marantz (New York, NY) portable solid-state recorder (PMD 671).

After a visual inspection of the recorded tokens, one clean audiorecording of *DEsert* and one clean audiorecording of *desSERT* were selected as the basis for manipulation. The duration, mean F0, mean intensity, and mean first three formants (F1, F2, F3) for each syllable in the selected *DEsert* and *deSSERT* are presented in Table II.

Based on the selected tokens, auditory stimuli were created that varied in seven steps from word-initial stress (step 1, *DEsert*) to word-final stress (step 7, *deSSERT*) along three dimensions: vowel quality, pitch, and duration (intensity cues to stress were neutralized for all stimuli). Three $7 \times 7$ matrices were created where the stimuli varied in two of these three dimensions while the third dimension was kept constant: a vowel quality $\times$ pitch matrix (duration cues neutralized at step 4); a vowel quality $\times$ duration matrix (pitch cues neutralized at step 4); and a pitch $\times$ duration matrix (vowel quality cues neutralized at step 4). The values corresponding to step 1 and step 7 in each dimension were those obtained in the acoustic analyses of the naturally produced tokens (Table II), and the values corresponding to step 2 through step 6 were generated in equidistant steps based on the acoustic measurements of the naturally produced stimuli. The values in the first and second syllables covaried in equidistant steps from the word-initial stress values to the word-final stress values.

The acoustic manipulations were performed in PRAAT version 6.0.46 (Boersma and Weenink, 2019). The stimulus with word-initial stress (*DEsert*) was used as the base token because the resulting auditory stimuli were judged by three additional English listeners (who did not complete the cue-weighting task) as sounding more natural. Only the voiced portion of the syllables was manipulated. The syllables first had their formant structure (F1, F2, F3, and the corresponding bandwidths) manipulated, followed by their duration, followed by their pitch, followed by their intensity. For the formant structure manipulation, a PRAAT script was adapted from Matt Winn's original script (Winn, 2014). Individual voicing portions within each word were manipulated separately. Formant and F0 tracking were first visually confirmed. Then the F1, F2, and F3 values, together with their associated bandwidths, were manipulated according to linearly interpolated values between the two end points identified in the acoustic analyses (Table II). High frequency noise was retained for naturalness considerations. For the duration manipulation, a PRAAT script was adapted from Matt Winn's original script (Winn, 2016). The duration tier function was used to stretch or squeeze the existing selected portion of the signal according to calculated ratios. For the pitch manipulations, the F0 of the voiced portions of each syllable was extracted in 10 intervals over the syllable for each stress pattern, and these 10 measurements were used to interpolate the pitch contour over the syllable using the *Manipulate* function of PRAAT to replace the original pitch tier with the interpolated values. For intensity, the *Scale Intensity* function of PRAAT was used to adjust the intensity of each syllable so that it would have the average intensity of the corresponding syllable across the two stress patterns. After these manipulations, the syllables were concatenated to form complete words, and the words had their overall

TABLE II. Acoustic measurements of the selected naturally produced stimuli.

| | *DEsert* | | *deSSERT* | |
|---|---|---|---|---|
| | First syllable | Second syllable | First syllable | Second syllable |
| Duration (ms) | 147 | 216 | 109 | 268 |
| F0 (Hz) | 275 | 208 | 227 | 223 |
| Intensity (dB) | 74 | 70 | 70 | 71 |
| F1 (Hz) | 669 | 556 | 432 | 597 |
| F2 (Hz) | 1845 | 1835 | 1841 | 1703 |
| F3 (Hz) | 2860 | 2242 | 2943 | 2082 |

intensity normalized to 70 dB. Visual representations of the formant structure, duration, and pitch manipulations and the complete set of auditory stimuli can be found in the supplementary materials.[3]

The acoustic manipulations generated 147 auditory stimuli (3 matrices of $7 \times 7$ stimuli). The auditory stimuli were presented in three blocks, with each stimulus being heard once per block and three times over the course of the experiment (total: 441 trials). To ensure that the participants knew how to carry out the task, the experiment also included 12 practice stimuli, where lexical stress was canonically realized along all three dimensions—with vowel quality, pitch, and duration being either at step 1 (*DEsert*) or at step 7 (*desSERT*) and with intensity being neutralized. The complete experiment thus included a total of 453 trials.

### C. Procedures

Participants completed the stress perception experiment at a computer station in a quiet lab. The experiment was delivered with the E-Prime 3.0 software (Psychology Software Tools, 2016). In each trial, participants heard an auditory stimulus over headphones and were instructed to press the left arrow on the keyboard if they thought they heard *DEsert* and the right arrow if they thought they heard *deSSERT*. The word labels and corresponding arrows appeared on the screen at the offset of the auditory stimulus. The next trial began 1000 ms after participants entered their responses, with the labels and arrows disappearing as participants entered their responses. The experiment began with a practice session in which participants heard the auditory stimuli where lexical stress was canonically realized. All participants reached an accuracy of 75% or higher in the practice session. The main session followed the practice session, with the auditory stimuli being fully randomized within blocks. The complete experiment lasted approximately 20–25 min.

### D. Data analysis

Participants' proportion of *DEsert* selection was coded as 1 and *deSSERT* selection as 0. Logit mixed-effects models were conducted on participants' responses using the lme4 package in R (Baayen *et al.*, 2008). Separate analyses were run for each of the three stimulus matrices (vowel quality by pitch, vowel quality by duration, pitch by duration). For each analysis, models including the two manipulated dimensions (each centered), L1 (with the English group as baseline), and their interactions as fixed effects were backward fitted using the log-likelihood ratio test.[4] All models included participant and item as crossed random effects. We report the models with the best fit. Whenever an effect of cue interacted with L1, we releveled the L1 variable with Dutch listeners as baseline and reran the model on the same data to determine whether the simple effects were significant for Dutch listeners. When two cues showed a significant effect, the fixed-effect coefficients were compared to determine whether one cue had a stronger effect than the other cue.[5] To investigate whether L2 learners' sensitivity to acoustic cues is modulated by their proficiency in English, and to determine whether L2 learners' reliance on one cue co-varies with their reliance (or lack thereof) on another cue, additional analyses were conducted on L2 learners' responses in which acoustic cues were included as random slopes for the participant in the L2 learner model with the best fit. L2 learners' individual random-slope coefficients were then obtained for each cue, inverted for interpretability (the higher the proficiency, the higher the inverted value should be),[6] and correlated with L2 learners' LexTALE scores and with each other. To eliminate the effect of outliers, relationships were analyzed using Spearman correlations. Effects are considered significant at or below an alpha level of 0.05.

## III. RESULTS

### A. Vowel quality by pitch stimuli

Figure 1 presents participants' responses when the stimuli varied by vowel quality and pitch. The left panel shows English listeners' results, and the right panel shows Dutch listeners' results. The $x$ axis represents the seven-step continuum of vowel quality (1 corresponding to word-initial stress and 7 corresponding to word-final stress); the $y$ axis

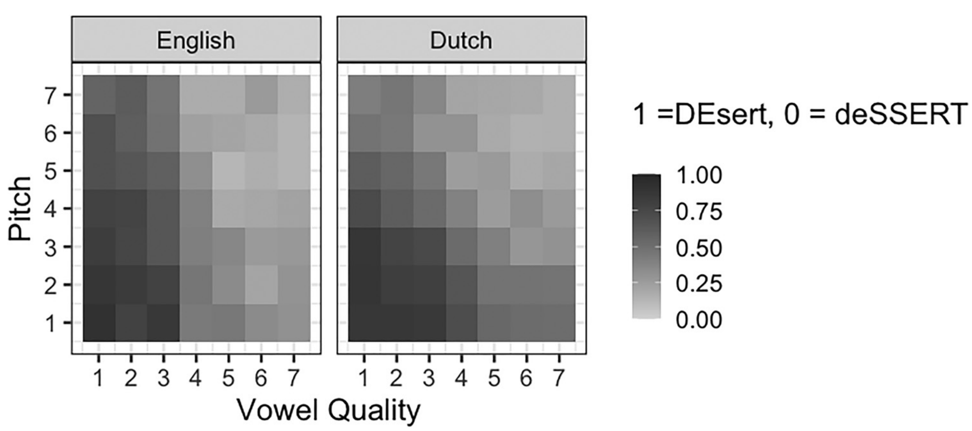


FIG. 1. Participants' proportions of *DEsert* vs *deSSERT* selection when the stimuli varied by vowel quality and pitch.

https://doi.org/10.1121/10.0005086



TABLE III. Results of logit mixed-effects model with best fit on participants' responses when the stimuli varied by vowel quality and pitch. SE, standard error.

| | Estimate | SE | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | –0.204 | 0.080 | –2.539 | 0.011 |
| Vowel quality | –0.517 | 0.027 | –19.270 | <0.001 |
| Pitch | –0.224 | 0.026 | –8.633 | <0.001 |
| L1 (Dutch) | 0.070 | 0.093 | 0.753 | >0.1 |
| Vowel quality × pitch | 0.026 | 0.011 | 2.347 | 0.019 |
| Vowel quality × L1 (Dutch) | 0.144 | 0.025 | 5.701 | <0.001 |
| Pitch × L1 (Dutch) | –0.156 | 0024 | –6.435 | <0.001 |

represents the seven-step continuum of pitch (1 corresponding to word-initial stress and 7 corresponding to word-final stress); and the color represents the proportion of *DEsert* vs *deSSERT* selection (dark gray for *DEsert* and light gray for *deSSERT*). Recall that these stimuli were at step 4 of the duration continuum, thus being halfway between the values expected for word-initial stress and the values expected for word-final stress. Table III summarizes the results of the mixed-effects model with the best fit on participants' responses when the stimuli varied by vowel quality and pitch.

The results in Table III revealed significant simple effects of vowel quality and pitch, with English listeners' proportions of *DEsert* responses decreasing as step number increases. A comparison of the fixed-effects coefficients for the model with English listeners as baseline revealed that the effect of vowel quality was stronger than that of pitch [$t(9849) = –8.269$, $p < 0.001$]. The mixed-effects model also yielded a significant two-way interaction between vowel quality and pitch, with the effect of vowel quality in English listeners' responses being greater at higher (i.e., more *deSSERT*-like) levels of pitch and with the effect of pitch being greater at higher (i.e., more *deSSERT*-like) levels of vowel quality. The fact that the three-way interaction did not improve the model [$\chi^2(1) = 0.0694$, $p > 0.1$] suggests that the two-way interaction between vowel quality and pitch is also true of Dutch listeners' responses (confirmed in the releveled model discussed below: estimate $= 0.026$, SE $= 0.011$, $z = 2.347$, $p < 0.019$). Crucially, the mixed-effects model revealed a significant two-way interaction between vowel quality and L1, with Dutch listeners showing a smaller effect of vowel quality compared to English listeners, and a significant interaction between vowel quality and L1, with Dutch listeners showing a greater effect of pitch compared to English listeners. To ascertain whether Dutch listeners showed a significant effect of vowel quality, we releveled the L1 variable with Dutch listeners as baseline. The model conducted on the same data with the releveled variable yielded significant effects of vowel quality (estimate $= –0.373$, SE $= 0.024$, $z = –15.702$, $p < 0.001$) and pitch (estimate $= –0.380$, SE $= 0.024$, $z = –15.967$, $p < 0.001$). A comparison of the fixed-effects coefficients for the model with Dutch listeners as baseline revealed that the effects of vowel quality and pitch did not differ ($t < |1|$, $p > 0.1$). These results indicate that while both English and Dutch listeners showed significant effects of vowel quality and pitch, English listeners evidenced a greater effect of vowel quality than of pitch and a greater effect of vowel quality than Dutch listeners, and Dutch listeners showed similar effects of vowel quality and pitch and a greater effect of pitch than English listeners.

Figure 2 represents the relationships between the effect of vowel quality and proficiency (LexTALE scores), the effect of pitch and proficiency, and the effects of vowel quality and pitch in individual L2 learners' data. The $x$ and $y$ axes represent the continuous variables (inverted slope of vowel quality effect, inverted slope of pitch effect, and LexTALE scores). The individual slopes were derived from a logit mixed-effects model conducted only on L2 learners' data, with vowel quality, pitch, and their interaction as fixed effects (model with the best fit). The more positive the value, the stronger the effect. A marginally significant positive Spearman correlation was obtained between the effect of vowel quality and proficiency (the higher the proficiency, the more sensitive to vowel quality) but not between the effect of pitch and proficiency, and no relationship was found between the effects of vowel quality and pitch. These results indicate that L2 learners became better able to use vowel quality as a cue to English lexical stress but did not

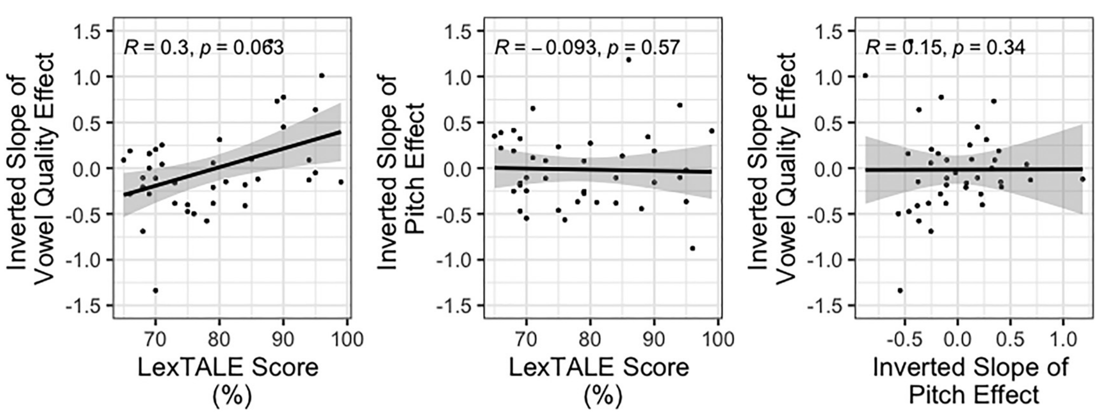


FIG. 2. L2 learners' relationships between the effect of vowel quality, the effect of pitch, and proficiency when the stimuli varied by vowel quality and pitch.

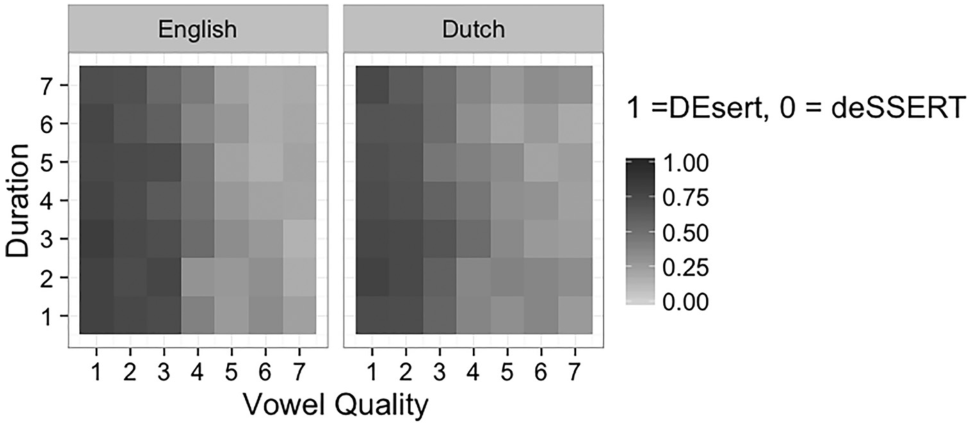


FIG. 3. Participants' proportions of *DEsert* vs *deSSERT* selection when the stimuli varied by vowel quality and duration.

decrease their reliance on pitch as they became more proficient in English, and L2 learners' use of vowel quality did not appear to depend on their use of pitch (or vice versa).

### B. Vowel quality by duration stimuli

Figure 3 presents English and Dutch listeners' responses when the stimuli varied by vowel quality and duration. This figure presents the same information as Fig. 1, but with the *y* axis representing the seven-step continuum of duration (1 corresponding to word-initial stress and 7 corresponding to word-final stress). These stimuli were at step 4 of the pitch continuum, thus being halfway between the values expected for word-initial stress and for word-final stress. Table IV summarizes the results of the mixed-effects model with the best fit on participants' responses when the stimuli varied by vowel quality and duration.

The results of the mixed-effects model presented in Table IV yielded significant effects of vowel quality and duration, with native English listeners' proportions of *DEsert* responses decreasing as step number increased. A comparison of the fixed-effects coefficients for the model with English listeners as baseline indicated that the effect of vowel quality was stronger than that of duration [$t(9849) = 15.041$, $p < 0.001$]. Importantly, the model revealed a significant two-way interaction between vowel quality and L1, with Dutch listeners showing a smaller effect of vowel quality compared to English listeners. To determine whether Dutch listeners showed a significant effect of vowel quality, we releveled the L1 variable with Dutch listeners as baseline. The model conducted on the same data with the releveled variable yielded significant effects of vowel quality (estimate = –0.390, SE = 0.021, $z$ = –18.793, $p <$ 0.001) and duration (estimate = –0.060, SE = 0.018, $z$ = –3.305, $p < 0.001$). A comparison of the fixed-effects coefficients for the model with Dutch listeners as baseline revealed that the effect of vowel quality was stronger than that of duration [$t(9849) = 12.025$, $p < 0.001$]. These results indicate that both groups showed significant effects of vowel quality and duration, and for both groups, vowel quality had a stronger effect on stress perception than duration, but English listeners evidenced a greater effect of vowel quality than Dutch listeners.

TABLE IV. Results of logit mixed-effects model with best fit on participants' responses when the stimuli varied by vowel quality and duration.

| | Estimate | SE | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | –0.151 | 0.088 | –1.711 | 0.087 |
| Vowel quality | –0.514 | 0.024 | –21.156 | <0.001 |
| Duration | –0.060 | 0.018 | –3.305 | <0.001 |
| L1 (Dutch) | –0.018 | 0.107 | –0.166 | >0.1 |
| Vowel quality × L1 (Dutch) | 0.123 | 0.025 | 5.019 | <0.001 |

Figure 4 represents the relationships between the effect of vowel quality and proficiency, the effect of duration and proficiency, and the effects of vowel quality and duration in L2 learners' data. The *x* and *y* axes represent the continuous variables (inverted slope of vowel quality effect, inverted slope of duration effect, and LexTALE scores). The slopes were derived from a logit mixed-effects model conducted only on L2 learners' data, with vowel quality and pitch as fixed effects (model with the best fit). A marginally significant positive Spearman correlation was obtained between the effect of vowel quality and proficiency (the higher the proficiency, the more sensitive to vowel quality) but not between the effect of duration and proficiency, and no relationship was found between the effects of vowel quality and duration. These results suggest that L2 learners became increasingly able to use vowel quality as a cue to English lexical stress but did not change their reliance on duration with increasing English proficiency, and L2 learners' use of vowel quality did not appear to depend on their use of duration (or vice versa).

### C. Pitch by duration stimuli

Figure 5 presents English and Dutch listeners' responses when the stimuli varied by pitch and duration. This figure presents the same information as in Fig. 3, but with the *x* axis representing the seven-step continuum of pitch. These stimuli were at step 4 of the vowel quality continuum, being halfway between the values expected for

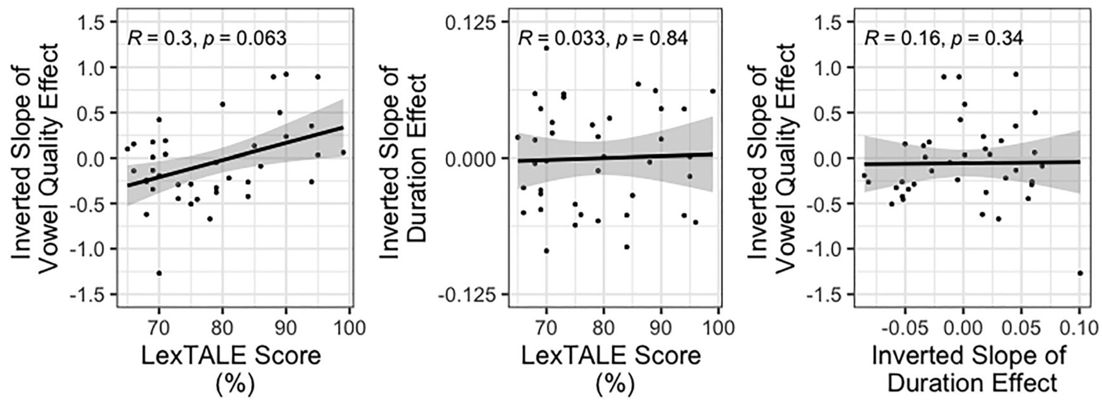


FIG. 4. L2 learners' relationships between the effect of vowel quality, the effect of duration, and proficiency when the stimuli varied by vowel quality and duration.

word-initial and word-final stress; this means that there was some vowel reduction in the first syllable of all stimuli. Table V summarizes the results of the mixed-effects model with the best fit on participants' responses when the stimuli varied by pitch and duration.

The results in Table V yielded significant effects of pitch and duration, with English listeners' proportions of *DEsert* responses decreasing as step number increases. A comparison of the fixed-effects coefficients for the model with English listeners as baseline revealed that the effect of pitch was stronger than that of duration [$t(9849) = -15.985$, $p < 0.001$]. Notably, the model also yielded a significant two-way interaction between pitch and L1, with Dutch listeners showing a stronger effect of pitch compared to English listeners, and a significant two-way interaction between duration and L1, with Dutch listeners showing a smaller effect of duration compared to English listeners. To establish whether Dutch listeners showed a significant effect of duration, we releveled the L1 variable with Dutch listeners as baseline. The model conducted on the same data with the releveled variable revealed a significant effect of pitch (estimate = –0.469, SE = 0.023, $z = -20.835$, $p < 0.001$) but no effect of duration (estimate = 0.036, SE = 0.022, $z = 1.628$, $p > 0.1$). These results indicate that while both English and Dutch listeners evidenced a significant effect of pitch, Dutch listeners showed a stronger effect of pitch than English listeners, and only English listeners showed a significant effect of duration.

Figure 6 represents the relationships between the effect of pitch and proficiency, the effect of duration and proficiency, and the effects of pitch and duration in L2 learners' data. The $x$ and $y$ axes represent the continuous variables (inverted slope of pitch effect, inverted slope of duration effect, and LexTALE scores). The slopes were derived from a logit mixed-effects model conducted only on L2 learners' data, with pitch as fixed effect (model with the best fit). No relationship was found between the effect of pitch and proficiency, but significant Spearman correlations were obtained between the effect of duration and proficiency (the higher the proficiency, the more sensitive to duration) and between the effects of pitch and duration (L2 learners who relied more on pitch also relied more on duration). The relationship between the use of pitch and duration remained significant even after the effect of proficiency was removed from the effect of duration ($r = 0.489$, $p = 0.002$).[7] These results indicate that L2 learners did not change their reliance on pitch as they became more proficient in English, but they increased their reliance on duration with increased proficiency despite the effect of duration not reaching significance for Dutch listeners in the group analysis, and those L2

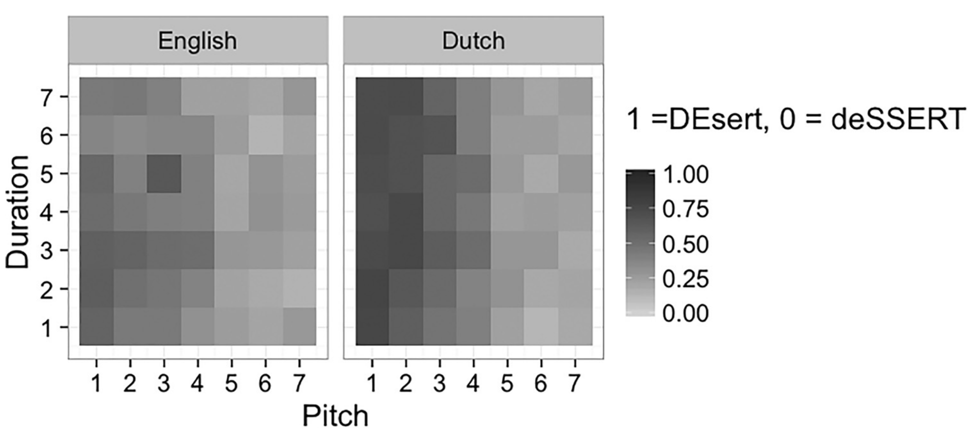


FIG. 5. Participants' proportions of *DEsert* vs *deSSERT* selection when the stimuli varied by pitch and duration.

TABLE V. Results of logit mixed-effects model with best fit on participants' responses when the stimuli varied by pitch and duration.

| | Estimate | SE | *z* | *p* |
|---|---|---|---|---|
| (Intercept) | –0.689 | 0.122 | –5.635 | <0.001 |
| Pitch | –0.261 | 0.024 | –10.803 | <0.001 |
| Duration | –0.048 | 0.024 | –2.008 | 0.045 |
| L1 (Dutch) | 0.404 | 0.153 | 2.643 | 0.008 |
| Pitch × L1 (Dutch) | –0.208 | 0.024 | –8.651 | <0.001 |
| Duration × L1 (Dutch) | 0.084 | 0.023 | 3.604 | <0.001 |

learners who relied more on pitch also relied more on duration (and vice versa) to perceive English lexical stress.

## IV. DISCUSSION

The present study used a cue-weighting stress perception task to investigate how English and Dutch listeners would weight acoustic cues to English lexical stress (RQ1), how L2 learners' cue weightings would change with increasing English proficiency (RQ2), and whether L2 learners' uses of individual cues to English lexical stress are interrelated (RQ3).

The results first showed that English listeners relied more heavily on vowel quality cues than on pitch or duration cues when perceiving English lexical stress, thus replicating the findings of previous studies (Chrabaszcz *et al.*, 2014; Zhang and Francis, 2010). The results also revealed that Dutch listeners as a group relied less strongly on vowel quality cues to English lexical stress than did English listeners, but their reliance increased as their proficiency in English improved. These results were true of both the stimuli where vowel quality and pitch had been manipulated and the stimuli where vowel quality and duration had been manipulated. These findings provide strong support for the cue-weighting transfer hypothesis: Vowel quality has been shown to be a weaker acoustic cue to Dutch lexical stress relative to suprasegmental cues such as duration (van Heuven and de Jonge, 2011). Dutch listeners' weaker use of vowel quality cues to English lexical stress compared to English listeners provides clear evidence that L2 learners transfer their weighting of acoustic cues from the L1 to the L2. Furthermore, Dutch listeners' increased sensitivity to vowel quality cues at a higher proficiency in English suggests that L2 learners *can* learn to show greater reliance on a cue that plays a more important role in the L2 than in the L1, in line with the findings of previous speech processing studies (Tremblay and Spinelli, 2014; Weber and Cutler, 2006). The finding that Dutch listeners weight vowel quality more strongly than duration (like English listeners and unlike what would be expected for Dutch; van Heuven and de Jonge, 2011) further suggests that some learning has taken place.

The results also indicated that Dutch listeners relied more strongly on pitch than did English listeners when perceiving English lexical stress, with their sensitivity to pitch not being modulated by their proficiency in English. Because previous studies on the perception of Dutch lexical stress (e.g., van Heuven and de Jonge, 2011) had not (to our knowledge) assessed the weight of pitch cues relative to segmental or other suprasegmental cues, Dutch listeners had been predicted to show stronger reliance on suprasegmental cues to English lexical stress compared to English listeners and compared to segmental cues, and stronger reliance on pitch than on duration given the intonational context in which the target words were elicited. The predicted L1 effect was confirmed for pitch, but Dutch listeners showed similar sensitivity to pitch and vowel quality. As discussed in the Introduction, because of the variety of F0 patterns that intonational pitch accents can take, pitch can be interpreted as a cue to lexical stress only in the context of a specific intonation for both Dutch (Gussenhoven, 2004; Krahmer and Swerts, 2001) and English (Beckman and Pierrehumbert, 1986; Ladd, 2012; Pierrehumbert, 1980; Pierrehumbert and Hirschberg, 1990). Thus, any difference between English and Dutch listeners' reliance on pitch cues to lexical stress is unlikely to stem from differences in the absolute weight of pitch cues in the two languages. However, and crucially, because vowel quality does not provide a strong cue to Dutch lexical stress, pitch has a greater *relative* weight to lexical stress in Dutch than in English. Dutch listeners' stronger reliance on pitch compared to

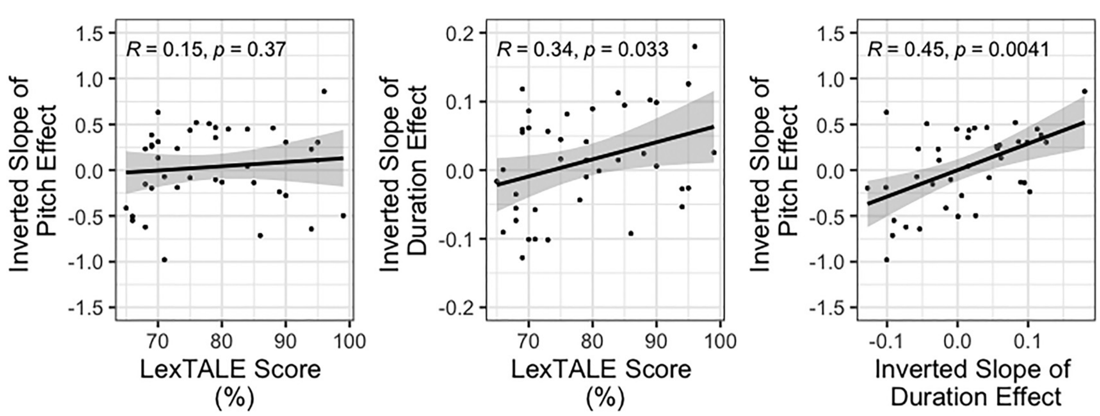


FIG. 6. L2 learners' relationship between the effect of pitch, the effect of duration, and proficiency when the stimuli varied by pitch and duration.

English listeners can therefore be attributed to this greater relative weight, in line with the cue-weighting transfer hypothesis. Importantly, unlike for vowel quality cues, Dutch listeners' reliance on pitch cues to English lexical stress was not found to be modulated by their English proficiency. Since English listeners evidenced less sensitivity to pitch cues than did Dutch listeners, one might have expected Dutch listeners to rely less on pitch cues to English lexical stress with increased proficiency in English, contrary to fact. This suggests that it may be difficult for L2 learners to suppress a cue that has a greater relative weight in the L1 than in the L2, corroborating the findings reported in previous speech processing research (Tremblay and Spinelli, 2014; Weber and Cutler, 2006). It is also possible that our Dutch listeners did not reduce their reliance on pitch with increased English proficiency because doing so was not detrimental to their perception of English lexical stress—it was in fact beneficial in an intonational context where the target words received a nuclear pitch accent. Further research is needed to determine whether the suppression of L1 cues is consistently more challenging than the learning of L2 cues.

The results further revealed that Dutch listeners showed a similar reliance on duration cues as English listeners when the stimuli had been manipulated for vowel quality and duration, and no reliance on duration cues when the stimuli had been manipulated for pitch and duration. Given the findings of van Heuven and de Jonge (2011), one would have perhaps expected Dutch listeners to rely more strongly on duration cues to English lexical stress than English listeners would have. Instead, L2 learners' weightings of vowel quality (in the stimuli manipulated for vowel quality and duration) and pitch (in the stimuli manipulated for pitch and duration) were so much stronger than their weighting of duration that they either surpassed (former) or completely overrode (latter) the effect of duration cues. One possible explanation of these results is that the attenuation of duration cues in sentence-final position, where the target words were elicited, exacerbated Dutch listeners' greater reliance on vowel quality and pitch, with the effect of duration disappearing when the other manipulated cue is one that Dutch listeners excel in using (pitch). This could explain the asymmetrical findings between the current study and that of van Heuven and de Jonge (2011), whose target words did not contain pitch cues and did not undergo phrase-final lengthening, potentially leading Dutch listeners to rely more on duration in their study (see also Sluijter and van Heuven, 1995, 1996b) than in ours [for similar results in Spanish and Catalan, see Ortega-Llebaria *et al.* (2010) and Ortega-Llebaria and Prieto (2011)]. Our finding that Dutch listeners rely more strongly on duration cues with increased English proficiency would be in line with an account where Dutch listeners learn to improve their reliance on duration cues to English lexical stress when these cues are weaker due to the intonational realization of the target words. All in all, although the current results for duration cues do not provide clear evidence in support of the cue-weighting transfer hypothesis, they do not refute it either and raise questions for further research at the intersection of lexical stress and intonation to investigate. Our findings, if replicable, would also suggest that Dutch listeners' greater sensitivity to suprasegmental cues to English lexical stress reported in previous studies (Cooper *et al.*, 2002; Cutler *et al.*, 2007) may have been driven by pitch cues rather than by duration cues.

Last but not least, the results indicated that Dutch listeners' use of one cue did not depend on their use of another cue, with the exception of pitch and duration, with Dutch listeners who showed greater sensitivity to pitch also showing greater sensitivity to duration. Thus, while L2 learners as a group can show cue-trading relationships (e.g., their reduced ability to use vowel quality cues corresponds to a stronger reliance on pitch cues), it is not the case that individual learners who show strong reliance on one cue necessarily show weak reliance on another cue. These results differ from those of studies that reported negative relationships between L2 learners' reliance on different cues to the same segmental contrast (e.g., Flege *et al.*, 1997; Schertz *et al.*, 2015). In the present study, the absence of negative relationships among the different cues suggests that L2 learners perceive lexical stress in a multidimensional way and independently weight acoustic cues to lexical stress according to their perceived functional weight, as determined by the L1, the L2, and L2 learners' proficiency in the target language. The positive relationship between Dutch listeners' use of pitch and duration cues, on the other hand, is consistent with the idea that some L2 learners have greater perceptual acuity and show a better ability to perceive suprasegmental cues to lexical stress contrasts, as reported in other studies on the perception of segmental contrasts (e.g., Clayards, 2018; Hazan and Rosen, 1991; Shultz *et al.*, 2012). Since a significant positive relationship between the use of acoustic cues was found in only one of three analyses, this finding should be interpreted with caution until it can be replicated in further research.

## V. CONCLUSION

The present study investigated whether Dutch L2 learners of English would transfer their weighting of acoustic cues to lexical stress from the L1 (Dutch) to the L2 (English). The results of a cue-weighting stress perception task in English showed that Dutch listeners relied less on vowel quality cues and more on pitch cues than English listeners, with their use of vowel quality increasing with English proficiency but with their use of pitch not being modulated by their English proficiency. These findings provide unequivocal support for a cue-based transfer approach to the perception of prosodic aspects of speech, with listeners' perception of lexical stress in the L2 being strongly influenced by the relative weight of acoustic cues to stress in the L1 (RQ1). These findings additionally suggest that it may be easier for listeners to increase their reliance on acoustic cues that have a greater functional weight in the L2 than in the L1 than it is to suppress their reliance on acoustic cues that have a weaker functional weight in the L2 than in the L1 (RQ2). The results further revealed that Dutch listeners' use of duration cues was not greater than that of native listeners, and it improved with increasing English proficiency when the stimuli had been manipulated for pitch and duration. These findings raise important questions on the relative weight of duration cues for further research on the perception of lexical stress to address. Finally, the results showed a positive relationship between L2 learners' use of pitch and duration cues, suggesting that some learners may be more skilled at perceiving suprasegmental contrasts than others (RQ3).

## ACKNOWLEDGMENTS

We are thankful to the audiences of the Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019 and the 17th Conference of the Association for Laboratory Phonology and to Jessamyn Schertz and another (anonymous) reviewer for their valuable comments on this research and manuscript.

[1]These studies did not specify how the stimuli were elicited from an intonational perspective. Because the words had pitch cues to lexical stress, they must have received a pitch accent, but it is unclear whether the words underwent phrase-final lengthening.

[2]LexTALE was developed to provide a quick and valid measure of lexical proficiency in advanced L2 learners of English. For Dutch L2 learners of English, LexTALE scores have been shown to correlate highly ($r = 0.86$) with general proficiency scores such as those obtained from the Quick Placement Test, an English proficiency test used within the Common European Framework for language levels (Lemhöfer and Broersma, 2012). We inferred from this relationship that LexTALE can be adopted with Dutch speakers as a measure of general English proficiency. Using a test that is not auditory in nature also decreases the likelihood of circularity between the proficiency test and the main focus of the study (lexical stress).

[3]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0005086 for the F1, F2, and F3, duration, and F0 values of each syllable for each manipulated step.

[4]The auditory stimuli were manipulated two acoustic cues at a time. For this reason, it was not possible to run a single model that included all the stimuli and all the effects in the analysis. With a single model, we would have needed to add stimulus matrix (vowel quality × pitch, vowel quality × duration, pitch × duration) as a fixed effect in the model. There is no clear theoretical basis for such an effect, and interactions with other fixed effects would be difficult to interpret given that different pairs of cues were manipulated in different stimuli.

[5]We used the following function to do so. Many thanks to Dr. Stephen Politzer-Ahles for suggesting that we use this function.

```
comparecoef <- function( refcoef, comparecoef, model ){
#variance-covariance matrix for all the coefficients in the model
varcov <- vcov(model)
#t statistic for the comparison of the two coefficients in the model
t <- (fixef(model)[comparecoef] - fixef(model)[refcoef]) /
sqrt(varcov[comparecoef,comparecoef] + varcov[refcoef,refcoef]
- 2*varcov[comparecoef,refcoef])
#two-tailed p value of the t statistic
p <- 2 * ( 1 - pt( abs(t), summary(model)$devcomp$dims["N"]-2 ) )
return(p) }
```

[6]Since participants' responses were coded as 1 for DEsert and 0 for deSSERT, we expect that the more sensitive participants are to the acoustic cue manipulations (step 1 for DEsert and step 7 for deSSERT), the more negative the slope should be. Inverting the value of the slope thus provides more readily interpretable results, with the slope being predicted to be more positive as proficiency increases.

[7]This was done by obtaining L2 learners' residuals from a linear regression where proficiency is a predictor of the (inverted) effect of duration and correlating these residuals with L2 learners' (inverted) effect of pitch.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**(4), 390–412.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accents* (Foris, Dordrecht, Netherlands).

Beckman, M. E., and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, edited by P. A. Keating (Cambridge University, Cambridge, UK), pp. 7–33.

Beckman, M. E., and Pierrehumbert, J. (**1986**). "Intonational structure in English and Japanese," Phonology **3**, 255–310.

Beier, E. J., and Ferreira, F. (**2018**). "The temporal prediction of stress in speech and its relation to musical beat perception," Front. Psychol. **9**, 431.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.0.46) [computer program]," http://www.praat.org (Last viewed 5/21/2021).

Braun, B., Lemhöfer, K., and Cutler, A. (**2008**). "English word stress as produced by English and Dutch speakers: The role of segmental and suprasegmental differences," in *Proceedings of INTERSPEECH 2008: 9th Annual Conference of the International Speech Communication Association*, September 22–26, Brisbane, Australia.

Brown, M. E., Salverda, A. P., Dilley, L. C., and Tanenhaus, M. K. (**2015**). "Metrical expectations from preceding prosody influence perception of lexical stress," J. Exp. Psychol. Hum. Percept. Perform. **41**, 306–323.

Cambier-Langeveld, T. (**1997**). "The domain of final lengthening in the production of Dutch," Linguistics Netherlands **14**, 13–24.

Chrabaszcz, A., Winn, M., Lin, C. Y., and Idsardi, W. J. (**2014**). "Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers," J. Speech Lang. Hear. Res. **57**, 1468–1479.

Clayards, M. (**2018**). "Differences in cue weights for speech perception are correlated for individuals within and across contrasts," J. Acoust. Soc. Am. **144**(3), EL172–EL177.

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., and Tremblay, A. (**2018**). "English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study," Lang. Learn. **68**, 635–668.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang. Speech **45**, 207–228.

Cutler, A., and Fox, D. J. (**1977**). "On the role of sentence stress in sentence processing," Lang. Speech **20**, 1–10.

Cutler, A., Wales, R., Cooper, N., and Janssen, J. (**2007**). "Dutch listeners' use of suprasegmental cues to English stress," in *Proceedings of the XVIth International Congress for Phonetic Sciences*, August 6–10, Saarbrücken, Germany, pp. 1913-1916.

Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phon. **25**, 437–470.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–1680.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Gay, T. (**1978**). "Physiological and acoustic correlates of perceived stress," Lang. Speech **21**, 347–353.

Gussenhoven, C. (**2004**). "Transcription of Dutch intonation," in *The Phonology of Intonation and Phrasing*, edited by S. A. Jun (Oxford University, Oxford, UK), pp. 118–145.

Gussenhoven, C. (**2008**). "Vowel duration, syllable quantity and stress in Dutch," in *The Nature of the Word: Studies in Honor of Paul Kiparsky*, edited by K. Hanson and S. Inkelas (MIT, Cambridge, MA), pp. 181–198.

Halle, M., and Vergnaud, J.-R. (**1987**). *An Essay on Stress* (MIT, Cambridge, MA).

Hammond, M. (**1999**). *The Phonology of English* (Oxford University, Oxford, UK).

Hazan, V., and Rosen, S. (**1991**). "Individual variability in the perception of cues to place contrasts in initial stops," Percept. Psychophys. **49**(2), 187–200.
Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.
Ingvalson, E. M., Holt, L. L., and McClelland, J. L. (**2012**). "Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency?" Bilingualism **15**, 255–274.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.
Jesse, A., Poellmann, K., and Kong, Y. Y. (**2017**). "English listeners use suprasegmental cues to lexical stress early during spoken-word recognition," J. Speech Lang. Hear. Res. **60**(1), 190–198.
Kim, D., Clayards, M., and Goad, H. (**2018**). "A longitudinal study of individual differences in the acquisition of new vowel contrasts," J. Phon. **67**, 1–20.
Krahmer, E., and Swerts, M. (**2001**). "On the alleged existence of contrastive accents," Speech Commun. **34**(4), 391–405.
Ladd, D. R. (**2012**). *Intonational Phonology* (Cambridge University, Cambridge, UK).
Lemhöfer, K., and Broersma, M. (**2012**). "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," Behav. Res. Methods **44**, 325–343.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1173–1181.
Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (**1981**). "Prosodic aspects of American English speech rhythm," Phonetica **38**, 84–106.
Ortega-Llebaria, M., del Mar Vanrell, M., and Prieto, P. (**2010**). "Catalan speakers' perception of word stress in unaccented contexts," J. Acoust. Soc. Am. **127**, 462–471.
Ortega-Llebaria, M., Olson, D. J., and Tuninetti, A. (**2019**). "Explaining cross-language asymmetries in prosodic processing: The cue-driven window length hypothesis," Lang. Speech **62**, 701–736.
Ortega-Llebaria, M., and Prieto, P. (**2011**). "Acoustic correlates of stress in Central Catalan and Castilian Spanish," Lang. Speech **54**(1), 73–97.
Pierrehumbert, J. (**1980**). "The phonology and phonetics of English intonation," Ph.D. dissertation, MIT, Cambridge, MA.
Pierrehumbert, J., and Hirschberg, J. (**1990**). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, edited by P. Cohen, J. Morgan, and M. Pollack (MIT, Cambridge, MA), pp. 271–311.
Psychology Software Tools (**2016**). "E-Prime 3.0 [computer program]," https://www.pstnet.com (Last viewed 5/21/2021).
Reinisch, E., Jesse, A., and McQueen, J. M. (**2010**). "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately," Q. J. Exp. Psychol. **63**, 772–783.
Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phon. **52**, 183–204.
Shields, J. L., McHugh, A., and Martin, J. G. (**1974**). "Reaction time to phoneme targets as a function of rhythmic cues in continuous speech," J. Exp. Psychol. **102**, 250–255.
Shultz, A. A., Francis, A. L., and Llanos, F. (**2012**). "Differential cue weighting in perception and production of consonant voicing," J. Acoust. Soc. Am. **132**(2), EL95–EL101.
Sluijter, A. M. C., and van Heuven, V. J. (**1995**). "Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch," Phonetica **52**, 71–89.
Sluijter, A. M. C., and van Heuven, V. J. (**1996a**). "Acoustic correlates of linguistic stress and accent in Dutch and American English," in *Proceedings of Fourth International Congress on Spoken Language Processing*, October 3–6, Newark, DE.
Sluijter, A. M. C., and van Heuven, V. J. (**1996b**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Small, L. H., Simon, S. D., and Goldberg, J. S. (**1988**). "Lexical stress and lexical access: Homographs vs. nonhomographs," Percept. Psychophys. **44**, 272–280.
Tremblay, A. (**2008**). "Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English," Appl. Psycholinguist. **29**, 553–584.
Tremblay, A., and Spinelli, E. (**2014**). "English listeners' use of distributional and acoustic-phonetic cues to liaison in French: Evidence from eye movements," Lang. Speech **57**(3), 310–337.
Turk, A. E., and Shattuck-Hufnagel, S. (**2007**). "Multiple targets of phrase-final lengthening in American English words," J. Phon. **35**(4), 445–472.
van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. **58**, 251–273.
van Heuven, V. J., and de Jonge, M. (**2011**). "Spectral and temporal reduction as stress cues in Dutch," Phonetica **68**(3), 120–132.
van Leyden, K., and van Heuven, V. J. (**1996**). "Lexical stress and spoken word recognition: Dutch vs. English," Linguist. Netherlands **13**, 159–170.
Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**(1), 597–607.
Winn, M. (**2014**). "GUI-based wizard for creating realistic vowel formant continua from modified natural speech (version 37)," http://www.mattwinn.com/praat/Make_Formant_Continuum_v37.txt (Last viewed 5/21/2021).
Winn, M. (**2016**). "Praat script: Calculate continuum steps," http://www.mattwinn.com/praat/Make_Duration_Continuum.txt (Last viewed 5/21/2021).
Zhang, Y., and Francis, A. (**2010**). "The weighting of vowel quality in native and non-native listeners' perception of English lexical stress," J. Phon. **38**, 260–271.