JSLHR

Research Article

Acoustic Cues to Perception of Word Stress by English, Mandarin, and Russian Speakers

Anna Chrabaszcz,[a] Matthew Winn,[b] Candise Y. Lin,[c] and William J. Idsardi[a]

Purpose: This study investigated how listeners' native language affects their weighting of acoustic cues (such as vowel quality, pitch, duration, and intensity) in the perception of contrastive word stress.
Method: Native speakers (N = 45) of typologically diverse languages (English, Russian, and Mandarin) performed a stress identification task on nonce disyllabic words with fully crossed combinations of each of the 4 cues in both syllables.
Results: The results revealed that although the vowel quality cue was the strongest cue for all groups of listeners, pitch was the second strongest cue for the English and the Mandarin listeners but was virtually disregarded by the Russian listeners. Duration and intensity cues were used by the Russian listeners to a significantly greater extent compared with the English and Mandarin participants. Compared with when cues were noncontrastive across syllables, cues were stronger when they were in the iambic contour than when they were in the trochaic contour.
Conclusions: Although both English and Russian are stress languages and Mandarin is a tonal language, stress perception performance of the Mandarin listeners but not of the Russian listeners is more similar to that of the native English listeners, both in terms of weighting of the acoustic cues and the cues' relative strength in different word positions. The findings suggest that tuning of second-language prosodic perceptions is not entirely predictable by prosodic similarities across languages.

Key Words: prosody, word stress, acoustic cues, perception, language typology, English, Mandarin, Russian

Perception of speech sounds is heavily influenced by the sound characteristics of a listener's native language (L1). This observation supports the idea that the native phonological system, which is acquired very early in life (Werker & Tees, 1984), filters out properties of the speech signal that are not relevant for the L1 (Polivanov, 1931; Shcherba, 1939; Trubetzkoy, 1969). Such perceptual bias has been extensively documented in literature on second language (L2) acquisition. It has been repeatedly shown that non-native vowel and consonant contrasts that do not exist in L1 are difficult for adults to discriminate and acquire (Best, Hallé, Bohn, & Faber, 2003; Best & Tyler, 2007; Flege & MacKay, 2004; Kuhl, 1991). A well-known example is that Japanese listeners experience difficulty with discrimination of the English /r/–/l/ contrast because these consonants are perceived as variants of the same phoneme in Japanese (Goto, 1971; Miyawaki et al., 1975). Spanish-dominant Spanish-Catalan bilinguals show decreased perceptual sensitivity to the Catalan /e/–/ɛ/ contrast compared with Catalan-dominant bilinguals presumably because Spanish collapses both vowels into the single category (Bosch, Costa, & Sebastián-Gallés, 2000; Navarra, Sebastián-Gallés, & Soto-Faraco, 2005; Pallier, Bosch, & Sebastián-Gallés, 1997). Such differences in segmental phonology highlight potential challenges for adult learners of a second language.

Languages differ not only in their repertoire of segmental contrasts but also in their suprasegmental (prosodic) properties. Perception of prosodic contrasts (e.g., word stress, tones) is affected by L1 phonology in at least two different ways: the stress pattern in a word (or the type of tone in tonal languages) and the acoustic cues used to realize prosodic contrasts. Some languages (e.g., French, Finnish, Turkish, Hungarian) do not generally contrast lexical items by stress pattern within a word. Consequently, speakers of such languages show relatively poorer ability to discriminate words and nonwords that differ in stress pattern (as in ˈ**per**mit—perˈ**mit**) compared with speakers of more stress-flexible languages like Spanish, Mandarin, Russian, and English

[a]University of Maryland, College Park
[b]University of Madison—Wisconsin
[c]University of Southern California, Los Angeles
Correspondence to Anna Chrabaszcz: lav@umd.edu

Editor: Rhea Paul
Associate Editor: Robert Marshall
Received October 11, 2013
Revision received December 18, 2013
Accepted January 24, 2014
DOI: 10.1044/2014_JSLHR-L-13-0279

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

(Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; C. Y. Lin, Wang, Idsardi, & Xu, 2014; Lukyanchenko, Idsardi, & Jiang, 2011). This pattern for stress perception parallels the widely observed trend that speakers of nontonal languages have difficulties with the perception of lexical tones (Bluhme & Burr, 1971; Kiriloff, 1969; Wang, Spence, Jongman, & Sereno, 1999).

Little is known about how acoustic attributes of prosodic contrasts in L1 affect perception of L2 prosody. Evidence to date indicates that speakers are adept at using acoustic cues in L2 if the same types of cues are actively used for prosodic contrasts in L1. For example, native speakers of Vietnamese are able to perceive stress contrasts in English despite a markedly different prosodic structure, because Vietnamese phonology already contains some important elements involved in English stress contrast, such as pitch cues (Nguyen & Ingram, 2005; Nguyen, Ingram, & Pensalfini, 2008). However, the same studies showed that Vietnamese speakers' production of other nonnative elements, such as duration contrast and vowel reduction, is typically incomplete or absent. Zhang, Nissen, and Francis (2008) observed a similar pattern of stress production for Mandarin speakers: Mandarin speakers did not reduce vowels to a native-like degree and also tended to use much higher pitch register. The implementation of L1 elements to L2 speech may aid stress perception and/or production but is potentially problematic; Juffs (1990) suggested that word stress is realized by native Mandarin speakers of English as pitch movement, with nonstandard realization of stress on nearly every word, including function words.

Under normal listening conditions, listeners typically do not perceive only one cue to the exclusion of others; they attend to a variety of cues at the same time and make their stress judgments on the basis of the weighting and interaction between the cues (Flege & Bohn, 1989; Fry, 1958), including contextual cues (McMurray & Jongman, 2011). The present study is based on the assumption that perception of stress is the result of the interaction of different acoustic cues. On the basis of this assumption, this study examines how such interaction of perceptual cues to nonnative word stress is influenced by the prosodic properties of a listener's L1. Many questions arise from this line of investigation, including those relevant for linguistic theory and language education. For example, when multiple cues to stress are in conflict with each other, how are the cues prioritized? Do listeners' L1 stress cues transfer to the processing of less familiar stress contrasts in L2? In addition, how are the strengths of these cues governed by their relative position in a word?

Perceptual Cues to Word Stress in English

The principle auditory cues that correlate with word stress include pitch (fundamental frequency, F0), duration, intensity, and vowel quality (Bolinger, 1961; Fry, 1958; Lehiste, 1970). Stressed syllables tend to have higher pitch, greater intensity, longer duration, and full (nonreduced) vowel quality, compared with nonstressed syllables. In this article, we adopt the view proposed by Liberman (1975) that syllables are perceived as stressed only by virtue of their relationship with nonstressed syllables. That is, stress is a relative property arising from a relationship between two (or more) syllables, rather than a property inherent in a syllable itself.

It has been proposed that the primary cue for stress in English in both natural and synthesized speech is relative pitch prominence (Beckman, 1986; Bolinger, 1958; Fry, 1958; Morton & Jassem, 1965). Beckman (1986) specifically suggested that F0 contour outranks amplitude contour, duration, and spectral quality (henceforth vowel quality). Fry (1958) manipulated F0 of synthetic disyllabic words and showed that a pitch difference of 5 Hz (re: 97 Hz) is sufficient to influence stress perception.

Relative importance of cues to stress perception remains poorly understood, perhaps because of the confounding relationships between cues and also between cues and speaking situations. For example, pitch of stressed syllables may vary depending on whether the syllables also carry phrase-level accent (i.e., prominence in the larger scope of the utterance). Stress and phrase-level pitch accent are acoustically correlated but linguistically distinct dimensions (Okobi, 2006; Plag, Kunter, & Schramm, 2011; Sluijter & van Heuven, 1996). Stress corresponds to the degree of prominence within a word and can be lexicalized, whereas phrase-level accent is used more variably by the speakers of a language to place focus on words in a sentence or a phrase as determined by the communicative situation (e.g., Eady & Cooper, 1986). Understandably, stressed syllables of accented words receive more prominence than stressed syllables of unaccented words.

In view of this, Sluijter and van Heuven (1996) suggested that F0 is the least reliable acoustic correlate in comparison to duration, vowel quality, or intensity for perception of English lexical stress. Okobi (2006) claimed that relative syllable duration is a strong and robust cue for stress prominence. Listeners are able to make reliable stress judgments on the basis of synthetic syllable duration differences (Adams & Munro, 1978; Cutler & Darwin, 1981; Isenberg & Gay, 1978), although synthetic speech has been found to yield general overestimation of the role of duration cues in segments (Assmann & Katz, 2005; Nittrouer, 2005). It should be noted, however, that vowel duration and vowel quality are also variable across different word and phonetic environments (M. Chen, 1970; Peterson & Lehiste, 1960; Raphael, 1972). The role of duration in English stress is likely constrained by durational differences arising from consonant environment (e.g., voiced vs. voiceless) and vowel height (low vs. high), which collectively introduce durational differences on the order of up to 4:1 in monosyllables (House, 1961).

Stressed vowels are typically higher in intensity than unstressed vowels (Beckman, 1986; Fry, 1955; Lieberman, 1960). However, intensity differences appear to be less effective in signaling stress than duration and pitch differences (Mattys, 2000; Morton & Jassem, 1965; Rietveld & Koopmans-Van Beinum, 1987; van Heuven & Menert, 1996). Furthermore, similar to durational cues, phonetic environment can

also have an effect on intensity, affecting vowel intensity within a range of roughly 18 dB even while controlling for cross-talker variability (House & Fairbanks, 1953).

In American English, stress contrasts are enhanced segmentally in terms of vowel quality (Beckman & Edwards, 1994; Campbell & Beckman, 1997; Fry, 1965). Unstressed vowels are produced in a more centralized position in acoustic F1–F2 space, which results in a less distinct vowel quality (i.e., unstressed vowels become more like a schwa): in "attic," /æ/ preserves a full quality, but in "attack," it is reduced to a schwa. Stress-related vowel quality is especially apparent for the peripheral vowels /i/, /a/, and /u/. In stressed syllables, these vowels tend to retain the most distinct spectral quality, whereas in unstressed syllables they tend to undergo substantial reduction (Rosner & Pickering, 1994). There are exceptions to the rule, however. Some normally reduced word-initial syllables can tolerate a full vowel (e.g., "morality," "photographic"). Full vowels can frequently occur in unstressed syllables of compound words (e.g., "grandma") or loan words (e.g., "ballet"). Sometimes reduced vowels can alternate with full vowels depending on the change in word meaning or the morphological form of the word (e.g., "to separate," verb, /ˈsɛpəreɪt/ vs. "separate," adjective, /ˈsɛpərət/). Because of such variability with regard to vowel reduction, there is no consensus about the precise weight of vowel quality in relation to other stress cues (Howell, 1993; Zhang & Francis, 2010). For example, Fry (1965) argued that the effect of vowel quality is outweighed by fundamental frequency, duration, and intensity differences; Beckman (1986) noted that vowel quality is at least a stronger cue than intensity. A study by Howell (1993) suggested that vowel quality is secondary only to pitch. Thus, the presence and dynamic nature of multiple cues has led to disparate results and little consensus on perceptual prominence.

Present Study

Although acoustic descriptions of word stress are well documented, the role of the individual cues and the interactions between cues are not well understood, especially in the context of perceiving a foreign language. As with other phonetic perceptions, the contribution of cues appears to be interactive, as contrast in one cue (e.g., pitch) can compensate for the lack of contrast in another cue (e.g., duration)—but only a few studies have examined the interaction of multiple cues at the same time (Rosner & Pickering, 1994; Sluijter & van Heuven, 1996; Zhang & Francis, 2010). Second, it is not clear whether the findings from research on English are generalizable to other languages, because the relative weights of acoustic cues in stress perception are likely to be language specific (Beckman, 1986; Dogil & Williams, 1999; Morton & Jassem, 1965). For example, duration and intensity cues are stronger than pitch cues for speakers of Czech (Janota & Liljencrants, 1969), whereas speakers of English show the opposite pattern. Furthermore, a cue that is an important signal of stress in L1 may be used for a different purpose in L2. In the context of linguistic variability in prosodic patterns, such variation is not surprising and is likely related to the degree to which the acoustic cues are required for other parts of the listener's L1 phonology (e.g., the use of vowel duration to mark both vowel and consonant contrasts as well as stress; e.g., Kondaurova & Francis, 2008; Koreman, van Dommelen, Sikveland, Andreeva, & Barry, 2009).

Cross-linguistic studies of stress perception remain scarce, and no study has compared pitch, duration, intensity, and vowel quality simultaneously. The primary goal of the present study was to provide further insights on perceptual correlates of word stress by examining the interaction of these four perceptual cues and to clarify how these cues are used by speakers of languages with typologically different prosodic systems. Specifically, we compared the relative weighting of the four principle acoustic cues in stress perception by native speakers of English, Russian, and Mandarin. These languages were chosen because they are typologically diverse and yet all can use word-level prosody in a contrastive way. For example,

English: ˊpermit–perˊmit

Russian: мýка– мукá (torture–flour)

Mandarin: dōng55 xī55–dōng55 xi2 (东西, EastWest–something)

Each of these languages uses prosodic cues differently. For example, English syllables may receive primary stress (first syllable in *au*tumn) or secondary stress (first syllable in *au*tomatic), or they may be fully reduced (second syllable in aut*o*matic). Russian is also a stress language, but it arguably lacks the distinction between primary and secondary stress, and Russian learners of English have been shown to incorrectly reduce English vowels with secondary stress (Banzina, 2012). Mandarin is a tonal language, and the way acoustic cues are used to implement tones differs significantly from how they are used to implement stress.

In this study, we examined how these four acoustic cues to prosodic contrasts may be used differently depending on whether they take the form of trochaic (strong–weak) or iambic (weak–strong) contours across both syllables of a nonce disyllabic word. There is ample evidence in the literature demonstrating that even when cue levels are identical in both syllables of a word, English listeners tend to perceive stress on the first syllable (trochaic stress bias; Baker & Smith, 1976; Morton & Jassem, 1965; van Heuven & Menert, 1996). Although L1 was expected to have some influence on how listeners perceived acoustic input and used acoustic cues, it remained unclear whether they would apply L1 strategies, approximate patterns exhibited by native listeners, or exhibit a random behavior that conformed neither to the L1 nor to the L2.

Stress in Russian. Russian has mobile stress, which can appear on any syllable and any morpheme (e.g., roots and affixes) of the word. Theoretical studies have determined the default stress position as initial, based on the interaction of phonological rules and morphology in the stress system of Russian (Halle & Vergnaud, 1987; Idsardi, 1992; Melvold, 1989). Russian stress is strongly centered: Unstressed syllables are arranged around the stressed syllable (Kerek et al., 2009). The most important correlate of Russian stress is vowel

reduction, both quantitative and qualitative (Badanova, 2007; Bondarko, 1977, 1998; Jones & Ward, 1969; Kijak, 2009; Kodzasov & Krivnova, 2001; Kondaurova & Francis, 2008; Zlatoustova, 1953). Stressed vowels are generally about 1.5–2 times longer than unstressed vowels depending on the word position and speech rate (Bondarko, 1977). In addition, Russian stress has a strong tendency to centralize vowels in non-pre-tonic positions and partially centralize vowels in pre-tonic positions (Badanova, 2007; Kijak, 2009). Unstressed vowels undergo different types of reduction depending on how far away they are from the stressed syllable. A three-step vowel reduction system is used to account for this phenomenon (first proposed by Potebnja, 1865). Stressed vowels have a full quality and are the longest; unstressed vowels in pre-tonic (pre-stress) positions are shorter, whereas all other unstressed pre-tonic vowels undergo even more significant quantitative and qualitative reduction (e.g., "сарафа́н," /sərʏˈfan/, "a traditional Russian dress"). Some researchers argue that there exists even a four-step reduction of Russian vowels because post-tonic vowels are weaker than pre-tonic vowels (Bondarko, 1998).

The second important correlate of stress in Russian is intensity (Jones & Ward, 1969), although its role is not completely understood. Some researchers suggest that syllable intensity is confounded with the unequal loudness of Russian vowels and phrasal prosody (Kodzasov & Krivnova, 2001). Syllable pitch difference has a minimal role in cueing word-level stress in Russian (Kijak, 2009), perhaps because of its role in indicating phrase-level prominence (Badanova, 2007).

Stress in Mandarin. Mandarin is a tonal language that differs dramatically from English and Russian in its use of acoustic cues to signal prosodic contrasts, although it does exhibit some similarities to stress languages (Y. Chen & Xu, 2006; Zhang et al., 2008). Although it is generally agreed that the main acoustic correlate of tones in Mandarin is the direction of the F0 contour during the vowel (Gandour, 1978; Shih, 1988), duration and intensity have also been found to correlate with the identification of tones in speech stimuli in which F0 information was impoverished while the original duration and intensity differences were preserved (Fu, Zeng, Shannon, & Soli, 1998; Liu & Samuel, 2004; Whalen & Xu, 1992). It remains to be seen whether those secondary cues can also contribute to lexical stress perception.

According to Van der Hulst (1999), tones can serve a contrastive function in two ways: when different tones can occur in the same positions (e.g., high vs. low) or when a tone can be present or absent on a certain syllable in a word (full and light syllables). Full syllables carry one of the four lexical tones, and syllables with these tones are pronounced louder and have greater duration and amplitude than light syllables (Duanmu, 2007; M. Lin & Yan, 1980; T. Lin, 1985; T. Lin & Wang, 1984). In contrast, light syllables carry the neutral tone and show considerable vowel reduction (Chao, 1968) and weaker articulatory strength (Y. Chen & Xu, 2006). Duanmu (2007) equated full syllables in Mandarin to stressed syllables in English, and light syllables in Mandarin to unstressed syllables in English (see also Chao, 1968). However, unlike English, where lexical stress is fixed and can be predicted by the metrical rules, lexical stress in Mandarin varies sociolinguistically and idiosyncratically (Shen, 1993). Shen (1993) found that in natural speech, neither the variation in F0 nor the variation in intensity changed the judgment of stress significantly, whereas duration exerted relatively more influence on stress perception. In production of reiterant speech, Shen identified that the duration ratio between full (stressed) and light (unstressed) vowels of the same quality is approximately 3:2, and the intensity difference is nearly 8 dB. When listening to nonnative stress contrasts in English, native Mandarin speakers have been shown to use duration, pitch, and vowel reduction but not intensity (Zhang & Francis, 2010).

In summary, studies on word-level stress in English, Russian, and Mandarin have identified the relevant acoustic cues to perception of stress. Based on the reviewed literature, Table 1 compiles the hypothesized relative importance given to the four cues (pitch, intensity, duration, and vowel quality) in the three languages. The purpose of the present study was to quantify the effect of L1 stress typology on the use of these cues in the same speech materials. Of interest were potential cross-language differences in stress cue weighting that could be related to prosodic typology. We predicted that, when exposed to the same acoustic signals, speakers of different languages would attend to acoustic cues to stress perception in a different fashion. If Mandarin and Russian listeners transferred their native perception strategies of lexical stress to L2 English, they would primarily attend to duration cues, whereas English speakers would primarily rely on pitch cues. However, if Mandarin listeners treated English stress contrasts as tonal differences, F0 cue should be the strongest. Vowel quality was hypothesized to be the second strongest cue for all three language groups. Intensity was hypothesized to be the weakest cue for English and Mandarin listeners; we expected pitch differences to be neglected by Russian listeners.

Method

Participants

Three groups of participants took part in this study: 15 native English speakers (10 women; age range: 22–55, $M = 27.7$), 15 native Russian speakers (11 women; age range: 22–35, $M = 26.8$), and 15 native Mandarin speakers (7 women;

Table 1. Hypothesized hierarchy of stress cues in English, Mandarin, and Russian (1 = *most important,* 4 = *least important*), based on previous literature.

	Language		
Stress Cue	**English**	**Mandarin**	**Russian**
Vowel quality	2	2	2
Pitch	1	3	4
Intensity	4	4	3
Duration	3	1	1

age range: 21–30, $M = 24.2$). All of the native English speakers were monolingual speakers born and raised in the United States. Russian speakers were born in Russia and had been living in the United States for an average of 3.8 years at the time of testing. Mandarin speakers were born in mainland China and had lived in the United States for an average of 1.3 years at the time of testing. A few participants in the Mandarin group may have been familiar with some regional dialects, but all of them were native speakers of standard Mandarin. Most of the participants were students in a university in the United States in or near Washington, DC; others were recent college graduates and were working in the United States at the time of testing. Many of the Russian and Mandarin participants were very advanced users of English. The length of formal instruction in English for the Russian participants ranged from 5 to 15 years ($M = 9.3$) and for the Mandarin participants from 4 to 16 years ($M = 9.6$). Before the experiment, all test takers were asked to fill out a language background questionnaire and a cloze test (a 50-item sentence completion task that assessed their English proficiency) developed by Brown (1980). This test was used as part of the English Language Institute placement test at the University of Hawai'i and is considered to be a valid and reliable measure of L2 learners' vocabulary, morpho-syntactic knowledge, and discourse competence. Native English speakers on average scored 95% correct, followed by Russian speakers (90% correct) and Mandarin speakers (75% correct). All participants reported no difficulty with hearing or speech. Informed consent was obtained for each participant, and the experimental protocol was approved by the institutional review board at the University of Maryland. Participants were reimbursed for their participation.

Materials and Design

The stimuli in this study were modified natural recordings of the disyllabic nonword "maba" produced by a phonetically trained male native speaker of North American English. Recordings of this word with stress on the first (/ˈmɑbə/) or second syllable (/məˈbɑ/) were chosen for clearest vowel quality and selected for further acoustic manipulation. This nonword was selected because it is phonologically and phonotactically permissible in Russian, English, and Mandarin, although the vowels in the recorded tokens were a schwa and a typical North American English low-back unrounded /ɑ/ vowel, which is qualitatively different than the low-central /a/ vowel in Russian and Mandarin.

Intensity, vowel quality, vowel duration, and pitch contour in both syllables were all manipulated independently using Praat (Boersma & Weenink, 2010) and Adobe Audition. Vowel quality combinations (/mɑbɑ/ and /məbə/) were created by splicing and pasting each syllable appropriately, resulting in four types of tokens: /mɑbɑ/, /məbə/, /mɑbə/, and /məbɑ/. Pitch and duration parameters were manipulated by resynthesis, and intensity was modified via waveform multiplication; all final sound stimuli were ultimately the result of Pitch Synchronous Overlap and Add (PSOLA) resynthesis. Acoustic values for strong and weak cue levels were proportional adaptations of the English stress production data reported by Zhang and Francis (2010; see Table 2). "Strong" levels included vowel /ɑ/, greater intensity, longer duration, and higher pitch, whereas "weak" levels included vowel /ə/, weaker intensity, shorter duration, and lower pitch. The current study used a 2 (pitch) × 2 (duration) × 2 (intensity) × 2 (vowel quality) factorial manipulation on each syllable. The fully crossed combinations of each of the four cues in both syllables resulted in 256 (i.e., 2^8) unique tokens. In many cases, there were conflicting cues. For example, a word could have a trochaic pitch contour (high–low pitch contour) but iambic vowel quality (/ə/–/ɑ/).

Procedure

Participants completed a forced-choice auditory identification task implemented through the Alvin software package (Hillenbrand & Gayvert, 2005), in which they were asked to identify the location of stress (first or second syllable stress) in a disyllabic nonword "maba" by clicking the corresponding buttons on a computer screen (Figure 1). Within each selection of "first" or "second" syllable stress, there were three buttons corresponding to the listener's degree of confidence in the choice. Thus, with the same button press, participants could select the placement of stress in the word and also provide a confidence rating of their decision on a 3-point scale. Stimuli were heard four times each; four 256-item blocks were presented, with randomized items in each block. In addition, participants received 16 practice trials before the actual experiment to familiarize them with the procedure and to make sure they understood the instructions correctly. The practice trials contained only unambiguous tokens where all four of the cues were contrastive across the two syllables and were not included in the final analysis. The average running time for the experiment was 50–60 min.

Analysis and Results

Depending on the combination of stress cues in each stimulus token, each cue could indicate contrastive stress on the first syllable ("strong"–"weak": trochaic) or the second syllable ("weak"–"strong": iambic) or be noncontrastive across syllables ("strong"–"strong": spondaic; or "weak"–"weak": pyrrhic). Stress cues were contrast coded such that each cue assumed a value of either maximally trochaic (1) or maximally iambic (–1). A zero value was assigned to the cue if it was noncontrastive (i.e., spondaic or pyrrhic) across the two syllables. For example, if pitch, duration, intensity, and vowel quality all indicated trochaic stress, the net trochaic contrast code would be 4. When the cues indicated an iambic stress, tokens were coded as –4. If three cues indicated trochaic stress but the fourth cue was noncontrastive, the code would be 3. If three cues indicated a trochee and the fourth cue indicated an iamb, the code would be 2. Clearly, several combinations of cues could have resulted in the token receiving the same code (e.g., 1 was assigned when any one of the four cues had a strong level in the first syllable while all

Table 2. Values of acoustic cues for the syllables "ma" and "ba" in "maba" in stressed and unstressed conditions.

Syllable and level	Vowel	Mean F0 (Hz)	Intensity (dB SPL)	Duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
Ma							
strong	ɑ	85	72	210	715	1200	2450
weak	ə	77.5	67	165	600	1250	2400
Ba							
strong	ɑ	84	72	200	730	1200	2400
weak	ə	77	67	150	550	1260	2400
Ratio (strong:weak)							
	ɑ	1.10	1.78	1.30	1.19	0.96	1.02
	ə	1.09	1.78	1.33	1.33	0.96	1.00
Ratio reported by Zhang et al. (2008)		1.10	1.78	1.35	N/A	N/A	N/A

Note. Cue levels are not equal across syllables, consistent with the recordings of natural utterances. The perceptual task thus measures not psychoacoustic sensitivity but sensitivity to the cue levels as they are typically observed. Ratio of intensity = $10^{((dB_{Stressed} - dB_{Unstressed})/20)}$. Formant values represent averaged formant frequencies over vowel duration. Ratios reported by Zhang et al. (2008) include mean values for all stressed and unstressed syllables in that study. N/A = no formant values for non-nasalized /a/ vowels are available in that publication.

other cues had noncontrastive levels [1 + 0 + 0 + 0 = 1], when two of the cues had strong levels in the first syllable, one cue had a strong level in the second syllable, and one cue had a noncontrastive value [1 + 1 – 1 + 0 = 1), and so on).

First, we examined the effectiveness of multiple cues on listeners' stress perceptions. Figure 2 suggests a direct relationship between the net number of cues for trochaic stress and the corresponding likelihood of trochaic stress perception. Listeners in all three language groups showed an increasing tendency to perceive stress as trochaic or iambic as they were presented with an increasing number of cues for those respective stress patterns. It is interesting that the relationship appears to be nearly linear across the entire range for all three listener groups. When the cues were used noncontrastively or when they canceled each other out, listeners did not show a preference for trochaic or iambic stress patterns (stress identification is at chance).

Second, listeners' responses were averaged for tokens with trochaic stress and for those with iambic stress contours. "Strength" of a particular cue was estimated as the proportion of perceptions consistent with a cue contour minus the proportion consistent with the opposite contour. Figure 3 illustrates the strength of each cue, as well as the contributions of each type of contrastive contour. For example, across all levels of vowel quality, intensity, and duration, a contrast in the pitch contour across syllables that indicated trochaic stress ("T" in Figure 3) resulted in roughly 8% more trochaic perceptions compared with when the pitch was noncontrastive; a change from noncontrastive to iambic ("I" in Figure 3) resulted in roughly 20% more trochaic perceptions. Together, these changes sum to 28%, which is the full "strength" of the cue (the filled circle in Figure 3). Vowel quality appeared to be the strongest cue for all listeners regardless of their L1 (consistent with the report of Zhang et al., 2008). Pitch was a strong cue for the English and Mandarin groups but not for the Russian group, who actually demonstrated inverse use of this cue (trochaic pitch contours increased likelihood of

Figure 1. Screen layout for the experiment interface.

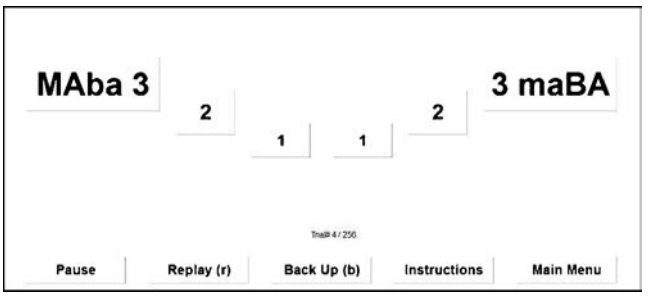


Figure 2. Cumulative effectiveness of cues for stress identification. A positive value means that the cues indicated a trochaic contour. A negative value means that the cues indicated an iambic contour. The *x*-axis reflects the cumulative sum of positive, negative, and/or neutral values of all four cues in relation to trochaic contours. Thus, 4 indicates that all four cues went in the trochaic direction, –4 indicates that all cues went in the iambic direction, 3 means that three of the four cues went in the trochaic direction, and so on.

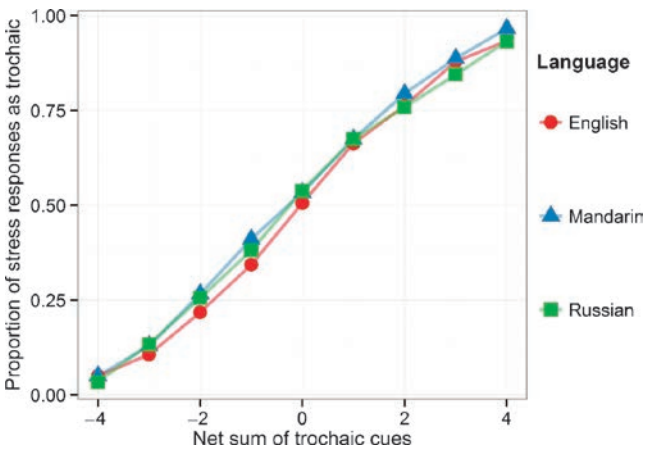
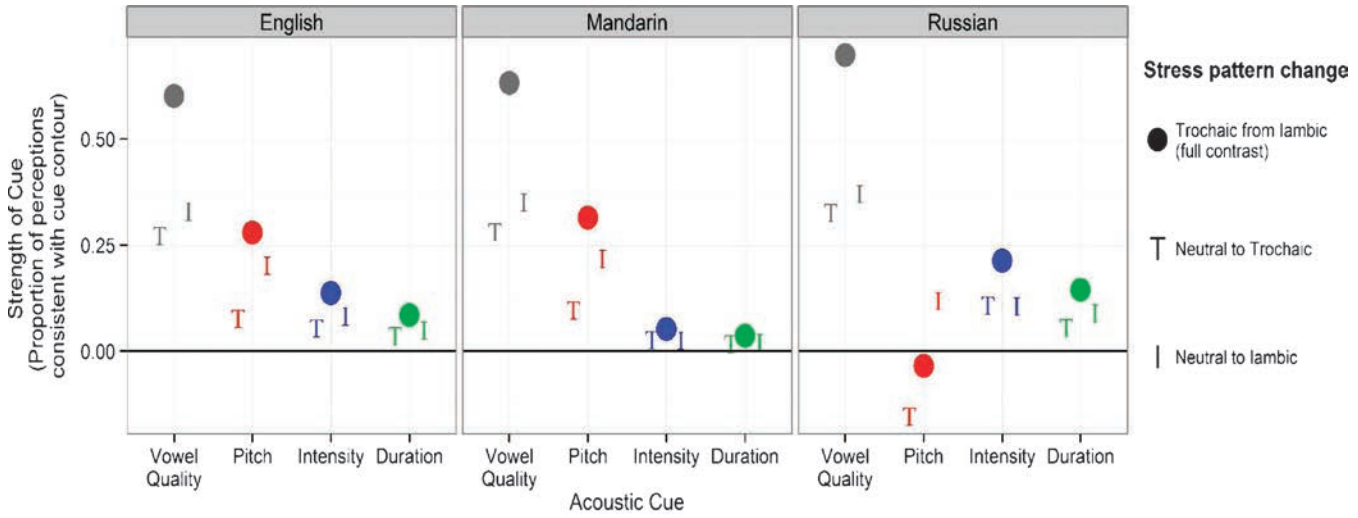


Figure 3. Strength of the acoustic cues in stress identification, arranged by language and stress pattern. "T" and "I" values reflect the increase in the proportion of perceptions as trochaic or iambic, respectively, when cue contours indicate corresponding stress patterns (minus the proportion when those cues were neutral or noncontrastive across syllables). In the case of the full-contrast measure (filled dots), the level represents the proportion of perceptions consistent with a cue contour minus the proportion consistent with the opposite contour. The negative value for trochaic pitch cues for the Russian group indicates that iambic perceptions were more common than trochaic perceptions, given trochaic pitch contours. Filled dots reflect the sum of trochaic and iambic subcomponents; values of the dots sum to 1.0 for each language group. Take the following example calculation: If a listener perceives noncontrastive pitch as half trochees and half iambs (50% trochees), and a trochaic pitch contour yields first-syllable stress responses 70% of the time, the strength of the trochaic contour is 0.20 (0.70 – 0.50). If the iambic pitch contour yielded 70% perception of iambs, the strength of the iambic contour is 0.20 (0.70 – 0.50). Iambic contours were generally stronger than trochaic contours.



iambic perceptions). The Russian group weighed intensity and duration to a greater extent than the other groups.

Figure 3 suggests that some cues are more influential for stress perception when they assume an iambic pattern; this means that, compared with the neutral or noncontrastive pattern, the iambic pattern is more perceptually distinct than the trochaic pattern. Overall, pitch and vowel quality were found to be generally stronger when in an iambic contour compared with when they were in a trochaic contour, and the other two cues were of roughly equal strength independent of stress pattern. This was not due to an overall trochaic bias in responses, as trochaic perceptions accounted for only 51%, 55%, and 55% of responses for the English, Russian, and Mandarin groups, respectively. Instead, this pattern could have arisen because of a bias to expect trochaic contours, in which case the iambic stimulus would be more clearly different than the expectation. Alternatively, it is also possible that the preference for iambic interpretations could be caused by the nature of the stimulus materials. The stimuli were recorded as single-word utterances, thus potentially eliciting utterance-level intonational influences (e.g., final lengthening of the second syllable "ba" compared with the first syllable "ma").

Finally, a statistical analysis of the listeners' responses was carried out. Listeners' binomial responses (trochaic or iambic stress choices) were analyzed using a generalized linear (logistic) mixed-effects model (GLMM) with a binomial family link function in the lme4 package (Bates & Maechler, 2010) of the R statistical computing software (R Core Team, 2013). Random effects were participants and items, and fixed effects included the four cues: vowel quality, vowel duration, pitch, and intensity. Native language of the listeners was also included as a fixed effect.

A series of fitted mixed-effects regression models were assembled and compared in order to find the optimal description of the data. Model comparisons were carried out using the Akaike information criterion (AIC) (Akaike, 1974), because it is recommended for the evaluation of mixed-effects models (Fang, 2011; Vaida & Blanchard, 2005). The basic goal of this criterion is to measure goodness of fit of the model without unnecessary parameter overfitting. A model with all cue factor terms, all two-way Cue × Language interactions, and a single three-way interaction of pitch, vowel quality, and language proved to be the most parsimonious fit to the data and reads as follows:[1]

> Stress ~ Pitch + Vowel Quality + Duration + Intensity + Language + Pitch:Language + Vowel Quality:Language + Duration:Language + Intensity:Language + Pitch:Vowel Quality:Language + (1 | Participant) + (1 | Item)

With such a design, the model's coefficients (factor estimates) should be interpreted as log odds of change in trochaic stress perception resulting from a change in the cue level from neutral to contrastive. For example, with an ideal

[1]Interaction between two factors is indicated by the colon; individual main effects are separated by plus signs.

intercept of 0 (50:50 odds), a listener presented with words with no contrastive cues will perceive half trochees and half iambs. A cue estimate of 1 represents a multiplicative change of *e* (approximately 2.72) in the odds ratio, say from 50:50 to 73:27; when that cue takes a trochaic contour, the listener will perceive a trochee 2.72 times more often compared with when it has a neutral value. The cues can be combined into an estimating equation such that any combination of cue contrasts can be modeled to predict perception. A negative interaction between main effects indicates subadditivity (i.e., the effect of one factor is reduced when the other factor changes in parallel). The coefficients for all of the main factors and interactions in the model are listed in Table 3.

The results of the GLMM analysis revealed that all four acoustic cues reached significance for each language group, and each group demonstrated a significantly different pattern of usage of each of the four cues ($p < .01$ for each comparison). Vowel quality was dominant for all language groups. Whereas the English and Mandarin groups weighed the pitch cue heavily, the Russian group demonstrated inverse use of this cue. The Russian listeners used intensity and duration to a greater extent than the English and Mandarin listeners. Mandarin-speaking listeners were more likely to identify syllables as trochaic even with no contrastive cues, but the trend did not reach significance ($p = .067$).

It is important to note that the four cues examined in this study are implemented in different domains (i.e., they have different units of measurement). Thus, although cross-group within-cue comparisons are incontrovertible, cross-cue comparisons of cue strength are inherently constrained by the degree to which different cue changes represent comparable changes in the stress domain. That is, a change from low pitch to high pitch is equal to a change from short to long duration only to the extent that these pitch and duration levels are externally valid. Insofar as much as the data derived from Zhang et al. (2008) are representative of the typical acoustic space of the stress contrast for all of these cues, we can provisionally assume some level of external validity. In addition, because the same cue levels are not equal across syllables (e.g., a "high" pitch in the first syllable is not the same F0 as a "high" pitch in the second syllable, consistent with the original natural utterances), the results reflect not psychoacoustic sensitivity but sensitivity to the cue levels as they are typically observed.

Table 3. Cue coefficients from the generalized linear (logistic) mixed-effects model (GLMM).

	Language		
GLMM coefficient	**English**	**Mandarin**	**Russian**
Intercept	–0.02	0.18	0.11
Vowel quality	1.58	1.70	1.91
Pitch	0.76	0.87	–0.07
Duration	0.23	0.10	0.42
Intensity	0.38	0.15	0.62
Pitch:Vowel quality	–0.01	0.05	–0.16
Total cue load	2.95	2.82	2.88

Discussion

The purpose of the present study was to examine how the use of four different acoustic cues to lexical stress in English—pitch, vowel quality, duration, and intensity—is affected by a listener's native language. The stress identification experiment used speakers of three typologically different languages: English, Russian, and Mandarin.

In contrast to our predicted hierarchy of stress cues for the three languages (see Table 1), the results of the GLMM analysis (Table 3) and summary of the raw data (Figure 3) indicated that there were more similarities between English- and Mandarin-speaking listeners than between English- and Russian-speaking listeners. The GLMM analysis suggested that for the English- and Mandarin-speaking listeners, all four cues were significant predictors of their stress perception, but pitch and vowel quality appeared to be the strongest cues, whereas duration and intensity cues were far less influential. Russian-speaking listeners demonstrated a distinctly different pattern. Although vowel quality was also the strongest perceptual cue for the Russian group, intensity and duration cues significantly contributed to stress identification performance, while the pitch cue was generally neglected. For the Russian group, all cues reached significance, but the pitch cue (the weakest cue) altered perception in the opposite direction. Table 4 presents the observed weighting of the four cues for each language group.

First, our results for the English-speaking listeners showed slight differences from our predicted cue hierarchy. Vowel quality was a stronger cue than pitch, and intensity was a stronger cue than duration; both of these patterns were opposite of what we had predicted.

With regard to Russian and Mandarin speakers, we predicted that, if L2 speakers transfer their L1 stress perception strategies to L2 tasks (e.g., Kondaurova & Francis, 2008; Nguyen & Ingram, 2005; Nguyen et al., 2008; Zhang et al., 2008), Russian and Mandarin L2 learners of English would primarily attend to those acoustic cues in the perception of English stress that are actively used for stress contrasts in their L1.

For the Russian group, the observed minimal role of pitch for English stress perception is consistent with previous studies on the perception of L1 stress (Badanova, 2007; Kijak, 2009), but the fact that vowel quality and intensity influenced stress perception to a greater extent than duration is somewhat surprising and is not consistent with the concept of L1 transfer. One possible explanation of this could be attributed to the fact that durational differences between unstressed and stressed syllables in English are on average smaller than those in Russian, which makes it more difficult for the Russian listeners to reliably identify stress on the basis of the English duration cue.

Consistent with the results of Zhang and Francis (2010), Mandarin-speaking listeners used vowel quality as a cue for stress identification. Mandarin vowels do not undergo qualitative reduction, and it has been previously shown that Mandarin-speaking speakers often have problems with vowel reduction in unstressed syllables in English productions

(Zhang et al., 2008). The results may arise from Mandarin-speaking listeners in our study treating the stressed /ɑ/ and the unstressed /ə/ in the disyllabic nonword "maba" as a segmental difference (i.e., as an English speaker might perceive /ɑ/ as different from /i/) rather than a stress-driven vowel quality difference (see Zhang and Francis [2010] for a more detailed discussion).

Among the four acoustic cues manipulated in this study, only vowel quality can be said to have a more stable connection with the lexical stress in terms of independence from other incidental speech characteristics (such as the overall loudness of speech, the rate of speech, and the emotional content of speech), which inevitably affect the differences between stressed and unstressed syllables in terms of intensity, duration, and pitch cues. These three cues are generally free to vary, and their utility for determining lexical stress is likely constrained to relative change across the word given a particular speech context (i.e., a low pitch is low only in the context of a higher pitch, while a schwa, as a rule, indicates an unstressed syllable regardless of its environment). This observation may help to explain the dominance of the vowel quality cue in all three groups of listeners; each syllable contains a vowel with an intrinsic cue to stress, whereas the other three cues must be derived from comparisons across syllables.

With regard to the effectiveness of different stress contours (iambic or trochaic), results of the current study suggest that for pitch and vowel quality, iambic patterns are more influential than trochaic patterns. That is, a low-to-high (or schwa to /ɑ/) sequence is more likely to change perception (from the neutral contour) than is the high-to-low (/ɑ/ to schwa) sequence. Duration and intensity cues did not yield this asymmetry, for reasons unknown; it should be noted, however, that these two cues were generally weaker than pitch and vowel quality.

This pattern may have resulted from the dominance of trochees in all three languages tested in the experiment; that is why a strong second syllable could be more salient because it is less common. In English, the stress pattern of 70% of disyllabic content words are trochaic (Cutler & Carter, 1987), which normally biases listeners to perceive stress on the first syllable even in words with exactly the same sound segments, identical pitch, intensity, and duration values in the two syllables (Morton & Jassem, 1965; van Heuven & Menert, 1996). In Russian, the default stress pattern is also word initial (Halle & Vergnaud, 1987; Idsardi, 1992), which could have accounted for the somewhat heightened stress sensitivity to the vowel cue in the iambic contour. Unstressed syllables cannot appear in word-initial positions in Mandarin (Duanmu, 2007), and vowel reduction can occur only in medial and final syllables. Hence, Mandarin listeners may have shown increased sensitivity to iambic contours, as seen by the other two language groups.

Table 4. Observed hierarchy of stress cues in English, Mandarin, and Russian (1 = *most important,* 4 = *least important*).

	Language		
Stress Cue	**English**	**Mandarin**	**Russian**
Vowel quality	1	1	1
Pitch	2	2	4
Intensity	3	3	2
Duration	4	4	3

Note. Despite similar weighting patterns, the results of the GLMM analysis revealed that each language group was statistically different from the others for the use of each of the four cues ($p < .01$ for each comparison).

In conclusion, there were similarities and differences in the cue-weighting patterns in the three listener groups. Of interest is the use of vowel quality by speakers of Mandarin, who do not implement vowel quality as a stress cue in their native language and demonstrate difficulty producing it in English stress contrasts (Zhang et al., 2008). It is possible that listeners in the Mandarin group had enough experience with English to recognize it as a cue, especially given the circumstances of the stimuli being produced by a native speaker of English. Among the notable differences between groups, Russian listeners showed negligible use of the pitch cue, especially when it indicated a trochaic contour. Russian listeners also demonstrated use of the intensity cue to a greater extent than the other two groups, albeit without the cue contour asymmetry that characterized other dominant cues.

In general, a comparison of stress perception performance of the speakers from three different language backgrounds offers some interesting insights into the cross-linguistic influence on prosodic processing. One of the most common concepts in second language acquisition literature is that larger linguistic differences between the speaker's L1 and L2 will result in higher probability of negative L1 transfer (interference). For example, when L1 and L2 prosodic elements differ substantially, L2 speakers may be more likely to use L2 prosodic cues incorrectly, which contributes to a detectable degree of foreign accent in production (Munro & Derwing 1995; Nguyen & Ingram, 2005; Nguyen et al., 2008; Trofimovitch & Baker, 2006; among others). In contrast, shared linguistic elements are believed to facilitate acquisition of similar elements in L2 (Corder, 1981; Kellerman, 1995; Ringbom, 1990, 2007). In this study, we observed that in spite of prominent differences in language typology, similar patterns of perception can arise. In addition, speakers of languages with similar prosodic elements can still demonstrate widely different patterns on prosodic perception tasks. Although English and Russian realize stress by means of different acoustic cues, both of them are stress languages. Mandarin is a tonal language that differs dramatically from both English and Russian in the use of certain cues but still maintains use of lexically contrastive stress. However, we see that stress perception performance of the Mandarin listeners but not the Russian listeners is more similar to that of the native English listeners, both in terms of weighting of the acoustic cues and their perceived strength in different word positions. This is even more notable because the Mandarin listeners scored significantly lower than the Russian listeners on the proficiency cloze test; yet they demonstrated a more native-like pattern of stress reliance.

The results of this study have some practical implications for language learning and teaching. Because prosodic properties facilitate language acquisition and serve as an essential part of the lexical code by which lexical entries are accessed and word boundaries are segmented (Cutler & Mehler, 1993), it is important to establish L1–L2 relationships at different levels of analysis, including prosody, to be able to predict which elements in L2 will present difficulties for learners of different L1 backgrounds. Our findings suggest that tuning of L2 prosodic perceptions may come not only through similarities but also through differences.

Acknowledgments

This study was supported by the University of Maryland's National Science Foundation–Integrative Graduate Education and Research Traineeship program in language (DGE-0801465, "Biological and Computational Foundations of Language Diversity").

References

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and nonnative speakers. *Phonetica, 35,* 125–156.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Assmann, P. F., & Katz, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *The Journal of the Acoustical Society of America, 117,* 886–895.

Badanova, T. A. (2007). *Word stress in the Altai language from the comparative perspective* (Unpublished doctoral dissertation). Gorno-Altaisk University, Gorno-Altaisk, Russia.

Baker, R., & Smith, P. (1976). A psycholinguistic study of English stress assignment rules. *Language and Speech, 19,* 9–27.

Banzina, E. (2012). *The role of secondary-stressed and unstressed-unreduced syllables in word recognition: Acoustic and perceptual studies with Russian learners of English* (Unpublished doctoral dissertation). Bowling Green State University, Bowling Green, OH.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes (R Package Version 0.999375-37) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, the Netherlands: Foris Publications.

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

Best, C. T., Hallé, P. A., Bohn, O.-S., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. *International Congress of Phonetic Sciences, 15,* 2889–2892.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience and second language speech learning: In honor of James Emil Flege* (pp. 12–34). Amsterdam, the Netherlands: John Benjamins.

Bluhme, S., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics, 22,* 51–57.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 4.5.16) [Computer program]. Retrieved from www.praat.org

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14,* 109–149.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language, 37,* 83–96.

Bondarko, L. (1977). *Sound system of the modern Russian language*. Moscow, Russia: Prosveschenie.

Bondarko, L. V. (1998). *Phonetics of the modern Russian language*. St. Petersburg, Russia: St. Petersburg University.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12,* 189–222.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal, 64,* 311–317.

Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Proceedings of the ESCA workshop on intonation: Theory, models and applications* (pp. 67–70). Athens, Greece: ESCA ETRW.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22,* 129–159.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech—evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica, 63,* 47–75.

Corder, S. P. (1981). Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics, 9,* 147–159.

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

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29,* 217–224.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics, 21,* 103–108.

Dogil, G., & Williams, B. (1999). The phonetic manifestation of word stress. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–311). Berlin, Germany: Mouton de Gruyter.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York, NY: Oxford University Press.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *The Journal of the Acoustical Society of America, 110,* 1606–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition, 106,* 682–706.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America, 80,* 402–415.

Fang, Y. (2011). Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. *Journal of Data Science, 9,* 15–21.

Flege, J. E., & Bohn, O. S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11,* 35–62.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1–34.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27,* 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152.

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. In X. Zwerner & W. Bethge (Eds.), *Proceedings of the 5th International Congress of Phonetics Sciences* (pp. 306–311). Munster: Karger/Basel.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America, 104,* 505–510.

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41–76). New York, NY: Academic Press.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia, 9,* 317–323.

Halle, M., & Vergnaud, J. R. (1987). Stress and the cycle. *Linguistic Inquiry, 18,* 45–84.

Hillenbrand, J. M., & Gayvert, R. T. (2005). Open source software for experimental design and control. *Journal of Speech, Language, and Hearing Research, 48,* 45–60.

House, A. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America, 33,* 1174–1178.

House, A., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America, 25,* 105–113.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America, 94,* 2063–2073.

Idsardi, W. J. (1992). *The computation of prosody* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Isenberg, D., & Gay, T. (1978). Acoustic correlates of perceived stress in an isolated synthetic disyllable. *The Journal of the Acoustical Society of America, 64,* S21.

Janota, P., & Liljencrants, J. (1969). The effect of fundamental frequency changes on the perception of stress by Czech listeners. *STL-QPSR, 4,* 32–38.

Jones, D., & Ward, D. (1969). *The phonetics of Russian*. Cambridge, England: Cambridge University Press.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics in Language Teaching, 28,* 99–118.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics, 15,* 125–150.

Kerek, E., Niemi, P., Parsons, C., Lyddy, F., Zhang, L., & Wu, A. (2009). Russian orthography and learning to read. *Reading in a Foreign Language, 21,* 1–21.

Kijak, A. (2009). *How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems* (Unpublished doctoral dissertation). Utrecht University, Utrecht, the Netherlands.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 20,* 63–67.

Kodzasov, S. V., & Krivnova, O. F. (2001). *General phonetics*. Moscow, Russia: RGGU.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America, 124,* 3959–3971.

Koreman, J., van Dommelen, W., Sikveland, R., Andreeva, B., & Barry, W. (2009). Cross-language differences in the production of phrasal prominence in Norwegian and German. In M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Proceedings of the Xth Conference of Nordic Prosody* (pp. 139–150). Frankfurt, Germany: Peter Lang.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Liberman, M. (1975). *The intonational system of English* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32,* 451–454.

Lin, C. Y., Wang, M., Idsardi, W., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*(2), 316–346.

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi [Acoustic properties of Mandarin neutral tone]. *Dialect, 3,* 166–178.

Lin, T. (1985). Preliminary experiments on the nature of Mandarin neutral tone [in Chinese]. In T. Lin & L. Wang (Eds.), *Working papers in experimental phonetics* (pp. 1–26). Beijing, China: Beijing University Press.

Lin, T., & Wang, W. (1984). Shengdiao ganzhi wenti [Perception of tones]. *Zhongguo Yuyan Xuebao* [*Bulletin of Chinese Linguistics*], *2,* 59–69.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when f0 information is neutralized. *Language and Speech, 47,* 109–138.

Lukyanchenko, A., Idsardi, W. J., & Jiang, N. (2011). Opening your ears: The role of L1 in processing of nonnative prosodic contrasts. In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the Second Language Research Forum 2010* (pp. 50–62). Somerville, MA: Cascadilla Press.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62,* 253–265.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118,* 219–246.

Melvold, J. (1989). *Structure and stress in Russian phonology* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8,* 159–181.

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38,* 289–306.

Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 912–918.

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

Nguyen, A.-T. T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics, 36,* 158–190.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *The Journal of the Acoustical Society of America, 118,* 1072–1088.

Okobi, A. (2006). *Acoustic correlates of word stress in American English* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64,* B9–B17.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32,* 693–703.

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics, 39,* 362–374.

Polivanov, E. (1931). La perception des sons d'une langue étrangère [Perception of sounds of a foreign language]. *Travaux du Cercle Linguistique de Prague, 4,* 79–96.

Potebnja, A. A. (1865). On phonetic peculiarities of Russian dialects. *Philological Notes, 1,* 63.

R Core Team. (2013). *R: A language and environment for statistical computing* [Computer program]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51,* 1296–1303.

Rietveld, A. C., & Koopmans-Van Beinum, F. J. (1987). Vowel reduction and stress. *Speech Communication, 6,* 217–229.

Ringbom, H. (1990). Effects of transfer in foreign language learning. In H. Dechert (Ed.), *Current trends in European second language acquisition research (Multilingual Matters Series)* (pp. 205–218). Clevedon, England: Multilingual Matters.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon, England: Multilingual Matters.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production* (Oxford Psychology Series No. 23). Oxford, England: Oxford University Press.

Shcherba, L. V. (1939). *Phonetics of the French language. A survey of French pronunciation in comparison with Russian: A manual for students*. Leningrad, Russia: Uchpedgiz.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech, 36,* 415–433.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory, 3,* 83–109.

Sluijter, A., & van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100,* 2471–2485.

Trofimovitch, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1–30.

Trubetzkoy, N. S. (1969). *Principles of phonology* (C. A. Baltaxe, Trans.). Berkeley: University of California Press. (Original work published in 1939).

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika, 92,* 351–370.

Van der Hulst, H. (1999). Word accent. In H. Van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 3–115). Berlin, Germany: Mouton de Gruyter.

van Heuven, V. J., & Menert, L. (1996). Why stress position bias? *The Journal of the Acoustical Society of America, 100,* 2439–2451.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106,* 3649–3658.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25–47.

Zhang, Y., & Francis, A. L. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38,* 260–271.

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,* 4498–4513.

Zlatoustova, L. V. (1953). *Phonetic nature of the Russian word stress* (Unpublished doctoral dissertation). Leningrad State University, Leningrad, Russia.