

Timo Roettger<sup>1</sup> / Matthew Gordon<sup>2</sup>

# Methodological issues in the study of word stress correlates

<sup>1</sup> University of Cologne, 50923 Cologne, Germany<sup>2</sup> University of California, Santa Barbara, CA 93106, USA, E-mail: mgordon@linguistics.ucsb.edu**Abstract:**

The investigation of acoustic correlates of word stress is a prominent area of research. The literature is rife with studies of the acoustic exponents of what is often referred to as stress but the methodological diversity of this research has created an unclear picture of the properties robustly associated with it. The present paper explores the methodological issues involved in examining word stress correlates with the goal of proposing a set of recommendations for future research. Based on a survey of 110 (sub-) studies on 75 languages, desiderata for research on the acoustics of stress are identified: descriptions of employed methods should be as detailed as possible, speech material should be designed to allow for teasing apart word level stress from phrase level prominence, and sample sizes should be chosen according to statistical considerations.

**Keywords:** word stress, phonetics, phonology, typology**DOI:** 10.1515/lingvan-2017-0006**Received:** March 7, 2017; **Accepted:** April 26, 2017

## 1 Introduction

The investigation of acoustic correlates of word stress is a prominent area of research dating back to pioneering work by Fry (1955, 1958). The literature is rife with studies of the acoustic exponents of stress but the methodological diversity of this research has created an unclear picture of the properties reliably associated with it. The present paper explores the methodological issues involved in examining word stress correlates with the goal of proposing a set of recommendations for future research.

## 2 Methodology

The present survey is gleaned from multiple sources, including papers found in phonetic and areal studies journals, working papers, and assorted books and dissertations. The study included only works containing quantitative results specifically targeting the acoustic correlates of stress (as opposed to acoustic correlates of other prosodic properties, e. g. intonation, prosodic constituency, boundary phenomena, rhythm and timing). Only studies involving adult speakers without reported speech impairments were considered.<sup>1</sup>

The corpus included a total of 110 (sub-)studies on 75 languages (or language varieties), which are plotted geographically in Figure 1 (see Gordon and Roettger this volume for a list of languages surveyed).

---

**Matthew Gordon** is the corresponding author.

© 2017 Walter de Gruyter GmbH, Berlin/Boston.



**Figure 1:** Geographical distribution of languages included in the survey of acoustic correlates of stress plotted via the “lingtypology” package (Moroz 2017) for R (2017).

For each surveyed source, besides the acoustic parameters signaling word stress (for discussion, see Gordon and Roettger this volume), we logged several methodological aspects.

The first was the corpus type examined, either lab speech or spontaneous speech (see Beckman 1997; Xu 2010). The former refers to speech that is usually recorded in the form of reading aloud pre-composed stimuli. Studies using lab speech examined target words in isolation, in a context phrase that varied depending on the target word, or in a fixed metalinguistic phrase, e. g. “Say \_\_\_\_\_ again” or “I say \_\_\_\_\_”. We also included larger pre-composed texts in this category. Spontaneous speech is here defined as “speech that is not read to script” (Beckman 1997: 7) and includes narratives, conversations, and interviews. In addition to the corpus type, the position of the target word is also provided, where options include phrase-final (including both final words in a phrase and words in isolation), phrase-non-final, or varied, in cases where the phrase position was not systematically varied. Cases in which final and non-final were crossed were labeled as ‘controlled’. If the position of the target word was unclear, a common situation for spontaneous speech, it was labeled as ‘unspecified’.

Also tracked was whether the target words occurred in a position that controlled for post-lexical tonal events such as accents. To operationalize encoding, five settings were adopted: accented, unaccented, accent controlled, implicit contrast or unspecified (in the case of spontaneous speech and studies employing carrier phrases that are not presented). The accented category includes contexts in which target words are either stated to or likely to carry a phrasal accent because they are explicitly focused, e. g. “I said \_\_\_\_\_ not *car*.” The unaccented category comprises cases in which the target word was either explicitly stated to be unaccented or in which focus fell on another word in the phrase, e. g. “I said \_\_\_\_\_ *slowly*, not *quickly*”. ‘Accented controlled’ reflects cases in which both the accented and unaccented conditions were crossed. Implicit contrast refers to cases in which focus was implicitly on the target word because it was the only varying element in the corpus, e. g. in a list of isolated words or words in an invariant (metalinguistic) phrase. It is impossible to know whether the target words in these cases carried a phrasal accent or not.

Additionally, we tracked the word stress levels examined (primary stress (1S), secondary stress (2S) and unstressed (US)) and the number of speakers, words (or alternative operationalizations for the corpus size), and repetitions of each targeted word under a given condition (applicable only for lab speech). For most studies, the most easily recoverable information was the number of words containing targeted syllables (or vowels). Depending on the corpus, each target word potentially consisted of one measured syllable (or vowel), either stressed or unstressed in studies employing cross-word comparisons, or more than one in intraword comparisons. For certain studies, the total number of data points rather than the number of targeted words is given since it was the only information provided.

The statistical test(s) employed and results for the examined acoustic parameters were also logged. Final notes columns provide additional information about the methodology and/or results. In coding statistics, descriptive studies are labeled as such and, in the case of those also using inferential statistics (which characteristically implies the simultaneous presentation of descriptive statistics), the type of test(s) employed is provided. Even though they are of interest, the statistical results of studies in our corpus are difficult to compare, since

there is considerable variation in the statistical models employed and the population over which inferences are drawn. Discussing the diversity of statistical choices and their potential impact on the reported results goes beyond the scope of this paper. Therefore, we have limited the present survey to encode the effect direction (e. g. greater duration of stressed syllables) of differences reported to be statistically significant regardless of the magnitude of the difference. Of course, given the diversity of statistical practices represented in the corpus, the reported significant differences should be regarded with caution. Furthermore, even differences that are genuinely statistically significant may not be perceptually relevant. In the absence of direct perception experiments, this distinction is difficult to assess, as even just-noticeable-differences reported in the literature and based on a small set of languages may not reflect difference limens for speakers of all languages. Although relying on statistical results to assess robustness runs the risk of attaching unwarranted importance to differences that may turn out not to be meaningful, we view this as a preferable alternative to dismissing results that could potentially be meaningful.

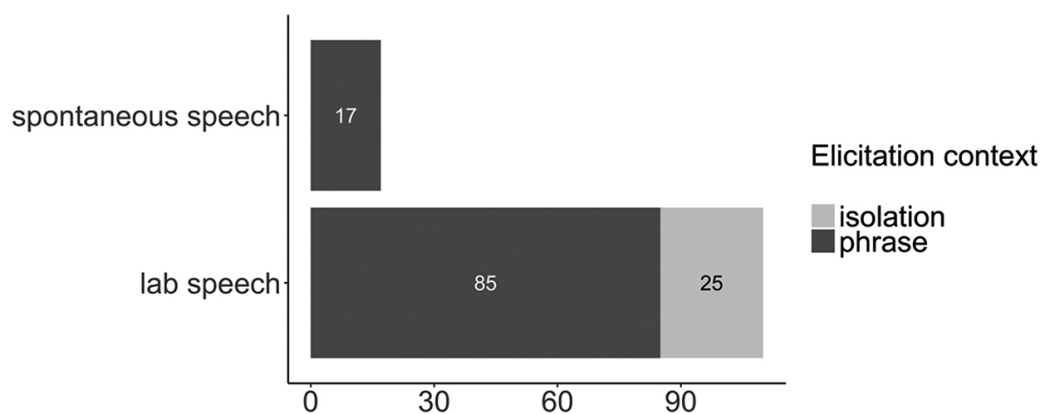
### 3 Results

#### 3.1 Teasing apart word-level stress from phrase-level prominence

One challenge in studying word stress is teasing apart word-level stress from properties attributed to other sources. A salient potential source of confound stems from phrase-level prominence (for a recent discussion, see Gordon 2014). In many languages, certain elements of the utterances are intonationally marked by tonal movements. In languages with word stress, pragmatically highlighted words often co-occur with post-lexical tonal events (pitch accents, phrasal accents, edge tones), which are usually realized on or near a syllable bearing word stress. Research conducted mainly on languages that exhibit pitch accents (e. g. English, French, Italian, and German) reports that the co-occurrence of an accent with a syllable results in temporal and spatial expansion of the articulatory gestures involved, which, in turn, leads to detectable acoustic differences (e. g. Cho 2005; Cho and Keating 2009; Cho and McQueen 2005; Harrington et al. 2000).

Languages also commonly modify temporal and spatial phonetic parameters at prosodic edges. There are two major phenomena falling under the umbrella of boundary-induced strengthening: “initial strengthening” and “final lengthening”. The former includes the phrase-initial enhancement of acoustic parameters that are phonemically contrastive. For instance, in languages with aspirated stops, voice onset time (VOT) is typically longer phrase-initially than phrase-medially (Cooper 1991; Pierrehumbert and Talkin 1992; Jun 1993; Cho and Jun 2000; Choi 2003; Cole et al. 2003). At the right edge, prosodic constituents are longer phrase-finally than phrase-medially in many languages, including Arabic (De Jong and Zawaydeh 1999), Dutch (Gussenhoven and Rietveld 1992), English (e. g. Beckman and Edwards 1990; Edwards et al. 1991; Wightman et al. 1992; Cho 2005; Turk and Shattuck-Hufnagel 2007), and Hebrew (Berkovits 1991) among many others.

These properties make it imperative to methodologically tease apart phonetic factors attributed to genuine word stress from those caused by phrasal phenomena such as prosodic boundaries and accentual prominence. Probably most crucial in avoiding these confounds is the phrasal context in which the target occurs. Figure 2 shows the number of studies in the database employing words elicited in lab speech (either in isolation or in a carrier phrase), and words elicited in spontaneous speech.



**Figure 2:** Number of studies as a function of corpus type (y-axis) and elicitation context in which target words appeared (color coded) for all (sub-)studies in the database. Note that studies using more than one type of context are included in the counts of all relevant categories.

As is evident in Figure 2, many studies in the database examined words in isolation, where word stress is straightforwardly confounded with phrase-level prominence. Studies based on spontaneous speech also run the risk of conflating word-level and phrase-level properties unless the target words are carefully chosen based on their role in the overall phrasal prosody. Many prominence asymmetries found in the survey thus might ultimately be attributed to phrase-level properties.

The majority of studies examine words in a phrase. In most of these, the target appeared in non-final position in at least one condition (which in some studies was crossed with final position). Riviera-Castillo and Pickering (2004) on Papiamentu, Astruc and Prieto (2006), Lehiste et al. (2005) on Meadow Mari, Lehiste et al. (2008) on Livonian, and Yakup and Sereno (2016) on Uyghur are exceptional in positioning the word only in phrase-final position, making it difficult to tease apart word stress from boundary-induced effects.<sup>2</sup>

There are several examples of potential interference from phrase-level effects in the database. Starting with the left-edge, Pycha (2006) finds that intensity in Turkish is *not* higher on a stressed final syllable than on an unstressed penult, contradicting results in Levi (2005). Pycha (2006) suggests that the lack of a difference in her study may be due to initial strengthening of the penult, which was also the first syllable in her database employing disyllabic words. Similarly, the lengthening of the initial stressed syllable in Estonian (Lehiste 1966) could be attributed to domain-initial strengthening.

Turning to the right edge, in their study of Indonesian stress in isolated words, Adisasmito-Smith and Cohn (1996) find that the vowel in the stressed penult for one of two speakers is longer than the pretonic but not the post-tonic (and final) vowel, the latter of which the authors suggest is lengthened due to its position, thereby obscuring a lengthening effect due to word stress. Cho (2006) observes that F<sub>0</sub> in Lakhota is higher on the stressed syllable of a disyllabic word when the stress is initial but not when the stress is on the second syllable. The absence of higher F<sub>0</sub> on a stressed second syllable is potentially due to final lowering, i. e. a phrase-final low boundary tone.

Similarly, higher F<sub>0</sub> on a stressed syllable may be attributed to post-lexical tonal events (e. g. Bolinger 1958, 1961; Beckman 1986; Ladd 2008 inter alia). Several studies, e. g. Gonzalez (1970) on Tagalog, Adisasmito-Smith and Cohn (1996) on Indonesian, Chian and Chiang (2005) on Saisiyat, Hargus and Beavert (2006) on Yakima Sahaptin, Gordon and Applebaum (2010) on Kabardian, and Simard et al. (2014) on Savosavo, find raised F<sub>0</sub> to be a correlate of stress in isolated words. It is unclear whether F<sub>0</sub> in these cases is attributed to word stress or to a post-lexical tonal event.<sup>3</sup>

Looking at lab speech stimuli embedded into a context phrase, we encounter difficulties in disentangling word prominence from phrasal prominence that is orthogonal to the position of the target word in the phrase. Intonation commonly expresses discourse relationships such as information structure (e. g. focus or topic) and information status (e. g. given vs. new information) (see, for example, Pierrehumbert 1980 on English, Grice et al. 2005a on Italian, Grice et al. 2005b on German, and Meyer and Mleinek 2006 on Russian). Many languages flag explicitly or implicitly focused constituents as well as new information with post-lexical tonal events which themselves might carry additional acoustic prominence. Thus, the speech material uttered in context phrases should carefully control for these discourse effects.

Vogel et al. (2016) demonstrate the potential influence of focus on the realization of prominence in a study of four languages (Spanish, Greek, Turkish, and Hungarian) that teases apart lexical prominence and phrasal prominence induced by focus through a series of dialogues. All target words in their study are non-final in a phrase but vary in whether they appear before another word that is narrowly focused or not. For example, the word *rope* is new focused information in the sentence “Ayse said ‘**rope**’ in the afternoon” uttered in response to the question “What did Ayse say in the afternoon?”. On the other hand, *rope* is background information that is not focused in the sentence “No. Ayse **said** ‘rope’ in the afternoon, she didn’t **write** it” in response to the question “Did Ayse write ‘rope’ in the afternoon?”. Vogel et al. (2016) find that stressed syllables in Hungarian have higher F<sub>0</sub> only when focused and that the effect of stress on duration and F<sub>0</sub> in stressed syllables is enhanced under focus in Greek. On the other hand, focus in Turkish induces a lowering effect on F<sub>0</sub> in the stressed final syllable relative to its non-focused counterpart. They suggest that this pattern reflects a phrase boundary following a focused word in Turkish. The Vogel et al. (2016) results demonstrate the interaction of focus with both post-lexical prominence and prosodic phrasing, both of which make the evaluation of the acoustic correlates of word stress difficult (see also, for example, Beckman and Pierrehumbert 1986 on Japanese, Kanerva 1990 on Chichewa, Hayes and Lahiri 1991 on Bengali, Jun 2005 on Korean).

Teasing apart phrasal prominence from word-level prominence is more subtle than merely ensuring the target word is not contrastively focused. Although metalinguistic carrier phrases like “I say \_\_\_\_\_ again” shield the target word from the right edge of the phrase, the targets in such phrases are invariably new information and implicitly contrasted with each other, making a post-lexical prominence on the word likely. In fact, in their study of English stress, Plag et al. (2011) classify the condition in which the target is embedded in the carrier phrase “She said \_\_\_\_\_ again” as the ‘accented’ condition, which they contrast with an ‘unaccented’ condition in



sentences with narrow focus on another word. The same context that is employed to elicit the focused condition in the Plag et al. study is thus the one used to trigger the non-focused condition in many other studies.

There are only seven studies in the database that employ a condition in which the target is clearly non-focused. These studies are not uniform, however, in the strategy they adopt to ensure the target is not focused; it is conceivable that the different defocusing constructions could induce their own prominence-enhancing or suppressing effects, e. g. post-focal compression (e. g. Bruce 1982 on Swedish, Cooper et al. 1985 on English, Féry and Kügler 2008 on German, Sadat-Tehrani 2008 on Persian and Tibetan, Lee and Xu 2010 on Korean, and Wang et al. 2011 on Uyghur).<sup>4</sup>

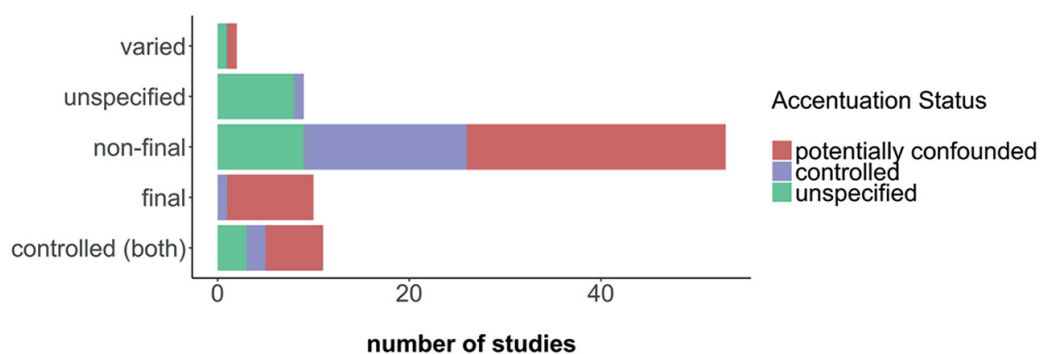
The potential conflation of post-lexical tonal events with word-level stress does not mean that studies that confound the two sources of prominence are completely uninformative. At least in languages that exhibit a post-lexical tonal event that is referred to as a 'pitch accent', the tonal event (per definition) docks on syllables carrying word stress. In these cases, prominence on a particular syllable in isolation is suggestive of primary word stress on that syllable.

In spite of the fact that post-lexical tonal events often co-occur with word stress in many well-described languages, one should nevertheless exercise caution when assuming a particular mapping of phrase-level prominence and word stress. Although tonal events such as pitch accents are typically assigned in "bottom-up" fashion docking on the primary stressed syllable of a word, there are many prosodic systems that exhibit tonal events that look like pitch accents but do not show the same temporal co-occurrence with stressed syllables. In Chickasaw, phrasal accent assignment operates orthogonally to word-level stress, leading to cases in question phrases in which a syllable that is unstressed at the word-level carries a tonal event (Gordon 2003). "Top-down" tonal placement in Chickasaw thus demonstrates the potential pitfalls of using phrasal accents to diagnose word stress. More generally, the link between phrasal accent and word stress commonly assumed relies on a particular interpretation of intonational events that may at least in part be an artifact of theoretical biases stemming from familiarity with more thoroughly studied (typically European) languages. Certain cases may be open to re-interpretation in different terms much like some "stress" languages, e. g. Indonesian (Goedemans and van Zanten 2007), have been re-analyzed as intonation-only languages.

Furthermore, even in languages in which phrasal accents diagnose word stress location, cues to word stress are not necessarily equivalent to those associated with phrasal accent. Word stress and phrasal accent may diverge acoustically, as shown in studies of several languages in which F<sub>0</sub> is a more reliable correlate of phrasal accent than word stress when phrasal position is controlled for, e. g. Huss (1978) on American English, Sluijter and Van Heuven (1996) on Dutch, Hintz (2006) on South Conchucos Quechua (2006), Ortega-Llebaria (2006) on Spanish, and Guion et al. (2010) on Balsas Nahuatl.

The upshot is that without detailed knowledge of how a language prosodically encodes discourse relations and the nature of its tonal events, the assessment of the source of acoustic prominence remains difficult.

Figure 3 illustrates the distribution of studies using lab speech according to the phrasal position of the target word and the accentual context. Most studies provided information on the target word position (all but 10). Most studies either controlled for phrase position by using stimuli with target words in both final and non-final positions (11) or used stimuli in non-final position (53). However, of these 64 studies, 12 studies did not provide any further information on potential confounds due to tonal events associated with information structure (green bars). Another 33 studies used either target words in accented positions or contexts in which the target words were explicitly contrasted (red bars). In these cases, the possibility that the observed prominence patterns are confounded with post-lexical prominence cannot be ruled out. In turn, this leaves us with 19 out of 85 studies (22%), that both controlled for phrasal position and assumed post-lexical prominence due to accent. Thus, a substantial number of studies in our survey are characterized by experimental design choices that make an interpretation of their results with regard to word stress difficult.

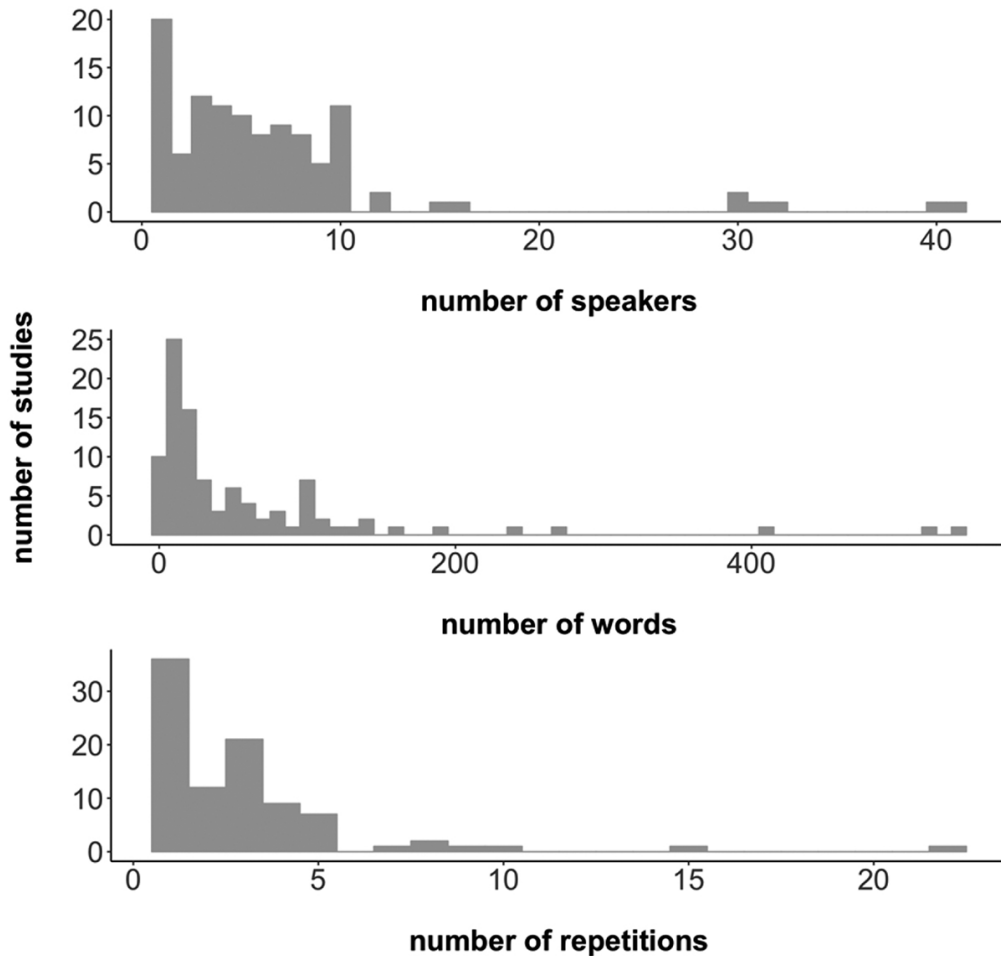


**Figure 3:** Number of studies as a function of target word position (y-axis) and accent information (color coded) within (sub-) studies using lab speech. Note that studies using more than one context are included in counts of all relevant categories.

### 3.2 Sample size

A crucial desideratum for any behavioral study, including those of acoustic correlates of stress, is the use of a large enough sample to confidently infer that the results reflect the broader population of speakers of a language beyond those contributing the data for the particular study. Additionally, a sufficient sample of lexical items is also important to ensure the generalizability of the obtained findings beyond the recorded words (e. g. Clark 1973). Another aspect of the design that affects the confidence with which we are able to estimate a representative effect is the number of recorded instances (referred to as repetitions) of a particular word produced by a particular speaker.

As Figure 4 shows, studies in the database vary considerably in the number of speakers, lexical items, and repetitions.



**Figure 4:** Number of speakers (top panel, bin width = 1), words (middle panel, bin width = 10), and repetitions (bottom panel, bin width = 1) represented in (sub-) studies in the database.

The majority of studies employ data from between one and ten speakers but vary widely within this range. On the other hand, the mode (20 studies) is a single speaker. The eight studies including data from more than 10 speakers range from 12 speakers in the study by Fry (1955) on American English and in the study by Everett (1998) on Pirahã to 41 in the work by Lesho (2013) on Cavite Chabacano.

It is interesting to note that only four studies reach the desired target of six speakers of each gender suggested by Ladefoged (1997: 140) in his recommendations for phonetic fieldwork. Roughly 46% (51 of 110) of the surveyed studies involve at least six speakers, which Ladefoged regards as the absolute minimum number necessary.

It should be emphasized, however, that Ladefoged’s recommendations are not based on statistical considerations. The minimum sample size allowing for statistical inference beyond the tested sample depends largely on statistical power, a dimension independent from significance (for a recent overview with regard to linguistic research, see Vasishth and Nicenboim 2016 or Kirby and Sonderegger). Assessing power is more difficult than assessing significance because it depends on multiple factors including the true (or expected) effect size, the sample size, and the degree of variability. For one, the effect size and its variability are strong determinants

of how large a sample should be to enable statistically robust inferences over a speaker population or a language's lexicon (or both). The variability of an effect, say a durational difference, is further dependent on the variability of the measurement. Speech data are very noisy and dependent upon many factors. Controlling for confounding factors is one way of reducing variability. Spontaneous speech is thus generally expected to yield more variable data due to the large number of confounding factors (higher level prosodic structure, intonation, syntax, etc.).

Another way to reduce variability and to achieve a better estimate of the effect is to increase observations for a single speaker. Researchers usually include many lexical items for a more precise estimate of the acoustic patterns of a particular speaker. This, in fact, might reduce variability attributed to the large variability across words, thereby enabling a better estimate of the behavior of the speech community. Looking at Figure 4, the majority of studies in our corpus use between 1 and 40 different lexical items. Depending on how a lexical item was defined (word type or word form) and the corpus (lab speech vs. spontaneous speech), some studies look at many hundreds of words.

Similarly, multiple repetitions of the same lexical item allow for a more precise estimate of the representative acoustic form of a specific word and thus reduces variability, again enabling a better estimate of the true effect in the population. Studies vary widely in repetition count ranging from one (where one repetition refers to only one instance of a word) in many studies to 22 in the Caldecott (2009) study of St'át'imcets. It is apparent in Figure 4 that the majority of studies use between 1 and 6 repetitions with a single repetition being most common (35 studies). An advantage to recording more than one token is that it allows for the possibility of discarding tokens associated with dysfluency or other localized problems with the recording. On the other hand, a large corpus consisting of many different lexical items ensures that the examined words are representative of those in the language.

### 3.3 Corpus composition

The issue of lexical breadth in a corpus is an important one that has received short shrift in the stress literature. In addressing the potential source of discrepancies between the results of their study of Tashlhiyt and those of Gordon and Nafi (2012), Roettger et al. (2015; see also Roettger accepted for a detailed analysis) suggest that a skewing in favor of words consisting of a light syllable followed by heavy syllable may have contributed to the evidence for final prominence in the Gordon and Nafi study. In a related vein, studies of Turkish (Levi 2005; Pycha 2006; Vogel et al. 2016) indicate a difference in the acoustic exponents of stress between words with exceptional penultimate stress vs. those with default final stress. These studies suggest that corpora should include sufficiently diverse word structures to uncover potential interactions between word type and the manifestation of prominence.<sup>5</sup>

Another related issue is whether comparisons between stressed and unstressed syllables are syntagmatic, i. e. involve comparisons within words, or paradigmatic, i. e. involve comparisons across words. The effects of a confounding variable potentially create differences between the two types of comparisons. For example, in their study of Hebrew, Silber-Varod et al. (2016) observe an interaction between position of the syllable and the realization of stress (see also Tuomainen et al. 1999 for a similar interaction in Finnish). Disyllabic words with final stress have higher F0 on the stressed final syllable, whereas disyllabic words with penultimate stress have *lower* F0 on the stressed penult. This result amounts to an effect of syllable position on F0 values such that the final syllable has higher values than the penult.<sup>6</sup> By comparing stressed and unstressed syllables across words with different stress positions, Silber-Varod et al. (2016) identified this effect, which would have otherwise escaped notice if their study had included only intraword comparisons.

## 4 Recommendations for best practice

In this section, we use the methodological variation observed in a survey of 110 (sub-) studies as a springboard for proposing recommendations for future research on word stress. These recommendations surround three key design aspects: the transparency of the experimental design, the speech material, and the sample size. For recommendations concerning instrumentation, e. g. microphones, recording equipment, file format, etc., the reader is referred to Ladefoged (1997, 2003), Bowerman (2008), Butcher (2013), Chelliah and De Reuse (2011), and Maddieson (2001).

First, the survey contains many studies lacking sufficient methodological details to allow for either evaluation or replication. Particularly crucial omissions in several studies include the phrasal position and intonational context in which target words were uttered. As a minimum requirement, future studies of stress should thus

make explicit (to the extent possible) the phrasal and accentual structure of their speech material (in addition, of course, to other relevant methodological details, e. g. corpus size, number of repetitions and speakers, etc.).

Second, it is imperative to control for context to ensure that the results reflect genuine word stress rather than phrasal prominence. Target words uttered in isolation make it difficult to disentangle word stress from phrase-level prominence. Moreover, target words should be removed from phrasal boundaries to avoid confounds attributed to either boundary-associated strengthening or tonal events. Finally, intonational structure should be controlled for, by either putting target words in an unaccented position or by examining both accented and unaccented positions. Only 19 lab speech studies in our corpus avoided these confounds altogether by controlling for phrasal and accentual structure. Another consideration is the comparability of the segments targeted for measurement since properties such as vowel quality, consonant type, surrounding sounds, and syllable structure all potentially impact common acoustic diagnostics of stress.

Third, a sample must be sufficiently large to infer that results reflect the broader population of speakers. A sufficient sample of lexical items is also important to ensure the generalizability of the obtained findings beyond the recorded words (e. g. Clark 1973; Winter 2011; Judd et al. 2012). The choice of the sample size is ultimately a question of statistical power (Vasishth and Nicenboim 2016 or Kirby and Sonderegger in press). Properties that determine statistical power are the (true) effect size, sample size, and degree of variability. Reduction of variability can be achieved by controlling for other factors interfering with measurements and by increasing the number of observations through additional words and/or repetitions. If the effect size and its variability can be estimated from previous studies, statistical power can be estimated in order to determine required sample size (see Kirby and Sonderegger in press). Consequently, even if some studies in our corpus remain anecdotal, i. e. do not allow for statistical inference over a broader population, they might allow us to generate hypotheses and represent valuable departure points for further investigations.

In summary, we hope that these recommendations based on evaluation of a survey of acoustic studies will encourage more careful experimental designs in future studies. Following these recommendations as much as possible given the constraints of a particular study will likely lead to more generalizable results within a language and to typological data that are easier to compare. It is hoped that this paper brings the relevance of experimental design to the attention of typologists, phoneticians, and phonologists, and that it encourages a renewed interest in experimental methods and statistical analyses.

## Acknowledgments

The authors thank Shigeto Kawahara, an anonymous reviewer, and the Department of Phonetics in Cologne for their helpful comments and feedback on an earlier version of this manuscript. Any remaining errors or misconceptions are our own.

## Notes

<sup>1</sup>The corpus (in form of a table) is publically available online at <https://osf.io/9r2cd/> alongside a script to reproduce respective counts presented in this manuscript. To establish a reliable and informative corpus that can be used in the future, cited authors are encouraged to submit corrections, if we have interpreted respective aspects of their method and/or results incorrectly. Further, we would like to invite scholars that have published work on word stress that is not logged in the present corpus to share their results with us for inclusion in the database.

<sup>2</sup>The controlled comparison of target words in Riviera-Castillo and Pickering (2004) involves words in final position, although they also investigate all syllables in another set of sentences. Astruc and Prieto (2006) examine accented and unaccented words in final position. The two Lehiste et al. studies (2005, 2008) target words in both phrase-final and sentence-final position.

<sup>3</sup>A similar interference of phrasal pitch accent also potentially obtains for words in final position of a phrase, a position toward which pitch accents tend to gravitate in phrases lacking narrow focus (Ladd 2008). There are relatively few studies in the database that do not explicitly control for accent but position words in phrase-final position. Among these, only Lehiste et al. (2005) on Meadow Mari and Lehiste et al. (2008) on Livonian find F0 to be a correlate of prominence.

<sup>4</sup>In two studies (Crosswhite 2003 on Bulgarian and Polish and Sadeghi 2011 on Persian), the target word appears in a carrier phrase in which another word is explicitly focused, e. g. "I say \_\_\_\_\_ again" where the target appears in prefocal position. The target in such phrases could still be implicitly focused because it is systematically varied while the phrase is held constant. These are categorized as deaccented to distinguish them from studies in which the carrier phrase lacks any explicit focus, a context in which the risk of implicit focus on the target is greater.

<sup>5</sup>If, however, speakers produce multiple items, the statistical analysis must account for this introduced non-independence (e. g. Clark 1973) to avoid statistically misguided interpretations (Judd et al. 2012; see also Winter 2011, for a brief discussion of this issue in phonetic experiments).

<sup>6</sup>Lieberman (1960), Crosswhite (2003), and Yakup and Sereno (2016) also report different results between intraword and interword comparisons of stressed and unstressed syllables.



## References

- Adisasmito-Smith, Niken and Abigail C. Cohn. 1996. Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working papers of the Cornell phonetics laboratory* 11. 1–16.
- Astruc, Lluïsa & Pilar Prieto. 2006. Acoustic cues of stress and accent in Catalan. *Proceedings of 3<sup>rd</sup> International Conference on Speech Prosody, Dresden, Germany*.
- Beckman, Mary. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
- Beckman, Mary. 1997. A typology of spontaneous speech. In Yoshinori Sagisaka, Nick Campbell & Norio Higuchi (eds.), *Computing prosody: Computational models for processing spontaneous speech*, 7–26. New York: Springer Science & Business Media.
- Beckman, Mary and Jan Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I*, 179–200. Cambridge: Cambridge University Press.
- Beckman, Mary & Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.
- Berkovits, Rochele. 1991. The effect of speaking rate on evidence for utterance-final lengthening. *Phonetica* 48. 57–66.
- Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14. 109–149.
- Bolinger, Dwight L. 1961. Contrastive accent and contrastive stress. *Language* 37(1). 83–96.
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. London: Palgrave Macmillan.
- Bruce, Gösta. 1982. Developing the Swedish intonation model. *Lund University, Dept. of Linguistics Working Papers* 22. 51–116.
- Butcher, Andy. 2013. Research methods in phonetic fieldwork. In Mark Jones & Rachael-Anne Knight (eds.), *The Bloomsbury companion to phonetics*, 57–78. New York: Bloomsbury Publishing.
- Caldecott, Marian. 2009. *Non-exhaustive parsing: Phonetic and phonological evidence from St'át'imcets*. University of British Columbia. Ph.D. dissertation.
- Chelliah, Shobhana & Willem De Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. New York: Springer.
- Chian, Wen-yu & Fang-mei Chiang. 2005. Saisiyat as a pitch accent language: Evidence from acoustic study of words. *Oceanic Linguistics* 44. 404–426.
- Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America* 117. 3867–3878.
- Cho, Taehong. 2006. An acoustic study of the stress and intonational system in Lakota: A preliminary report. *Speech Sciences* 13. 23–42. (Published by The Korean Association of Speech Sciences).
- Cho, Taehong and Sun-Ah Jun. 2000. Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society* 36. 31–44.
- Cho, Taehong and Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37. 466–485.
- Cho, Taehong and James McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33. 121–157.
- Choi, Hansook. 2003. Prosody-induced acoustic variation in English stop consonants. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain*, 2661–2664.
- Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12. 335–359.
- Cole, Jennifer, Hansook Choi, Heejin Kim & Mark Hasegawa-Johnson. 2003. The effect of accent on the acoustic cues to stop voicing in Radio News speech. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain*. 2665–2668.
- Cooper, A. 1991. Glottal gestures and aspiration in English. PhD dissertation Yale University.
- Cooper, William E., Stephen J. Eady & Pamela R Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77. 2142–2156.
- Crosswhite, Katherine. 2003. Spectral tilt as a cue to word stress in Polish, Macedonian, and Bulgarian. *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences, Barcelona, Spain*, 767–770.
- De Jong, Kenneth & Bushra Adnan Zawaydeh. 1999. Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics* 27. 3–22.
- Edwards, Jan, Mary Beckman & Janet Fletcher. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89. 369–382.
- Everett, Keren. 1998. The acoustic correlates of stress in Pirahã. *Journal of Amazonian Languages* 1(2). 104–162.
- Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36. 680–703.
- Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27. 765–768.
- Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1. 120–152.
- Goedemans, Rob & Ellen van Zanten. 2007. Stress and accent in Indonesian. *LOT Occasional series* 9. 35–62.
- Gonzalez, Andrew. 1970. Acoustic correlates of accent, rhythm, and intonation in Tagalog. *Phonetica* 22. 11–44.
- Gordon, Matthew. 2003. The phonology of pitch accent placement in Chickasaw. *Phonology* 20. 173–218.
- Gordon, Matthew. 2014. Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels. In Henry van der Hulst (eds.), *Word stress: Theoretical and typological issues*, 83–118. Cambridge: Cambridge University Press.
- Gordon, Matthew & Ayla Applebaum. 2010. Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40. 35–58.
- Gordon, Matthew & Latifa Nafi. 2012. The acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics* 40. 706–724.
- Gordon, Matthew & Timo B. Roettger. This issue. *Acoustic correlates of word stress: A cross-linguistic survey*.
- Grice, Martine, Stefan Baumann & Ralf Benz Müller. 2005b. German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (eds.), *Prosodic typology – The phonology of intonation and phrasing*, 55–83. New York: Oxford University Press.

- Grice, Martine, Mariapaola D'Imperio, Michelina Savino & Cinzia Avesani. 2005a. A strategy for intonation labelling varieties of Italian. In Sun-Ah Jun (eds.), *Prosodic typology – The phonology of intonation and phrasing*, 362–389. New York: Oxford University Press.
- Guion, Susan, Jonathan D. Amith, Christopher S. Doty & Irina A. Shport. 2010. Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language. *Journal of Phonetics* 38. 137–166.
- Gussenhoven, Carlos & A. C. M. Rietveld. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20. 283–303.
- Hargus, Sharon & Virginia Beavert. 2006. A note on the phonetic correlates of stress in Yakima Sahaptin. *University of Washington Working Papers in Linguistics* 24. 64–95.
- Harrington, Jonathan, Janet Fletcher & Mary Beckman. 2000. Manner and place conflicts in the articulation of accent in Australian English. In Michael B. Broe & Janet Pierrehumbert (eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon*, 40–51. Cambridge: Cambridge University Press.
- Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47–96.
- Hintz, Diane. 2006. Stress in South Conchucos Quechua: A phonetic and phonological study. *International Journal of American Linguistics* 72. 477–521.
- Huss, Volker. 1978. English word stress in the post-nuclear position. *Phonetica* 35. 86–105.
- Judd, Charles M., Jacob Westfall & David A. Kenny. 2012. Treating stimuli as random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology* 103. 54–69.
- Jun, Sun-Ah. 1993. The phonetics and phonology of Korean Prosody. PhD Dissertation Ohio State University.
- Jun, Sun-Ah. 2005. Korean intonational phonology and prosodic transcription. In Sun-Ah Jun (ed.), *Prosodic typology – The phonology of intonation and phrasing*, 201–229. New York: Oxford University Press.
- Kanerva, J. 1990. *Focus and phrasing in Chichewa phonology*. New York: Garland.
- Kirby, James & Morgan Sonderegger. in press. Model selection and phonological argumentation. In Diane Brentari & Jackson Lee (eds.), *Shaping phonology*. Chicago: University of Chicago Press.
- Ladd, D. Robert. 2008. *Intonational phonology*. New York: Cambridge University Press.
- Ladefoged, Peter. 1997. Instrumental techniques for phonetic fieldwork. In William Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*, 137–166. Malden, MA: Blackwell Publishing.
- Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell Publishing.
- Lee, Yong-cheol & Yi Xu. 2010. Phonetic realization of contrastive focus in Korean. *Proceedings of 5<sup>th</sup> International Conference on Speech Prosody, Chicago*.
- Lehiste, Ilse. 1966. *Consonant quantity and phonological units in Estonian*. Bloomington: Indiana University Press.
- Lehiste, Ilse, Pire Teras, Valts Ernštreits, Pärtel Lippus, Karl Pajusalu, Tuuli Tuisk & Tiit-Rein Viitso. 2008. *Livonian Prosody*. Helsinki: Suomalais-Ugrilainen Seura (*Mémoires de la Société Finno-Ougrienne* 255).
- Lehiste, Ilse, Pire Teras, Toomas Help, Pärtel Lippus, Einar Meister, Karl Pajusalu & Tiit-Rein Viitso. 2005. *Meadow Mari Prosody (Linguistica Uralic Supplementary Series 2)* Tallinn: Teaduste Akadeemia Kirjastus.
- Lesho, Marivic. 2013. The sociophonetics and phonology of the cavite chabacano vowel system The Ohio State University PhD dissertation.
- Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35. 73–97.
- Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32. 451–454.
- Maddieson, Ian. 2001. Phonetic fieldwork. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 211–229. New York: Cambridge.
- Meyer, Roland & Ine Mleinek. 2006. How prosody signals force and focus: A study of pitch accents in Russian yes–no questions. *Journal of Pragmatics* 38. 1615–1635.
- Moroz, George. <http://CRAN.R-project.org/package=lingtypology> 2017. *Lingtypology: Linguistic typology and mapping*.
- Ortega-Llebaria, Marta. 2006. Phonetic cues to stress and accent in Spanish. *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, 104–118.
- Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. PhD Dissertation MIT.
- Pierrehumbert, Janet and David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard Docherty & D. Robert Ladd (eds.), *Papers in laboratory phonology II: Gesture, segment, prosody*, 90–117. Cambridge: Cambridge University Press.
- Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39. 362–374.
- Pycha, Anne. 2006. A duration-based solution to the problem of stress realization in Turkish. UC Berkeley Phonology Lab Annual Reports.
- R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Riviera-Castillo, Yolanda & Lucy Pickering. 2004. Phonetic correlates of stress and tone in a mixed system. *Journal of Pidgin and Creole Languages* 19(2). 261–284.
- Roettger, Timo B. accepted. *Tonal placement in Tashlhiyt Berber – How an intonation system accommodates to adverse phonological environments*. Studies in Laboratory Phonology. Berlin: Language Science Press.
- Roettger, Timo B., Anna Bruggeman & Grice, Martine. 2015. Word stress in Tashlhiyt – Postlexical prominence in disguise. *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Hong Kong.
- Sadat-Tehrani, Nima. 2008. The structure of Persian intonation. *Proceedings of 4<sup>th</sup> International Conference on Speech Prosody, Campinas, Brazil*, 249–252.
- Sadeghi, Vahid. 2011. Acoustic correlates of lexical stress in Persian. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences, Hong Kong*, 1738–1741.
- Silber-Varod, Vered, Hagit Sagi & Noam Amir. 2016. The acoustic correlates of lexical stress in Israeli Hebrew. *Journal of Phonetics* 56. 1–14.
- Simard, Candide, Claudia Wegener, Albert Lee & Connor Youngberg. 2014. Savosavo word stress: A quantitative analysis. *Proceedings of 7<sup>th</sup> International Conference on Speech Prosody, Dublin, Ireland*.
- Sluijter, Agaath M. C. & Vincent J. Van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100. 2471–2485.

- Tuomainen, Jyrki, Stefan Werner, Jean Vroomen & Beatrice De Gelder. 1999. Fundamental frequency is an important acoustic cue to word boundaries in spoken Finnish. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences, San Francisco*, 921–923.
- Turk, Alice & Stefanie Shattuck-Hufnagel. 2007. Phrase-final lengthening in American English. *Journal of Phonetics* 35. 445–472.
- Vasishth, Shravan & Bruno Nicenboim. 2016. Statistical methods for linguistic research: Foundational ideas – part I. *Language and Linguistics Compass* 10(8). 349–369. DOI: 10.1111/lnc3.12201. DOI:10.1111/lnc3.12201.
- Vogel, Irene, Angeliki Athanasopoulou & Nadia Pinkus. 2016. Prominence, contrast, and the functional load hypothesis: An acoustic investigation. In Jeffrey Heinz, Rob Goedemans & Harry van der Hulst (eds.), *Dimensions of phonological stress*, 123–167. Cambridge: Cambridge University Press.
- Wang, Bei, Ling Wang & Tursun Kadir. 2011. Prosodic encoding of focus in six languages in China. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences, Hong Kong*.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 92. 1707–1717.
- Winter, Bodo. 2011. Pseudoreplication in phonetic research. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Science, Hong Kong*. 2137–2140.
- Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics* 38(3). 329–336.
- Yakup, Mahire & Joan Sereno. 2016. Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association* 46. 61–77.