Lexical Stress Matures Late in Typically Developing Children

Joanne Arciuli\textsuperscript{1}, Kate Phillips\textsuperscript{2}, Benjamin Bailey\textsuperscript{1}, Alexandre Forndran\textsuperscript{1}, Adam Vogel\textsuperscript{3}, Kirrie Ballard\textsuperscript{4}

\textsuperscript{1}Flinders University, Australia  
\textsuperscript{2}University of New South Wales, Australia  
\textsuperscript{3}University of Melbourne, Australia  
\textsuperscript{4}University of Sydney, Australia

Joanne.Arciuli@flinders.edu.au

Abstract

Previous studies on the development of lexical stress production in English include narrow age ranges. We addressed this issue by eliciting acoustic data from 253 typically developing 3-12 year olds and adults. Here we report a subset of results from our larger study. We report the results from 12 weak-strong (WS) words. WS words may be more challenging for English speaking children due to trochaic bias. We measured stress contrastivity (duration, intensity and fundamental frequency) across the first two syllables of each correct word production using the normalized pairwise variability index (PVI). While word productions were intelligible and exhibited stress contrastivity, generally, children did not exhibit adult-like magnitudes of stress contrastivity.

Index Terms: speech, speech production, lexical stress, acoustic analyses

1. Introduction

Most research on development of lexical stress production includes studies of English [1]-[6]; with a smaller number of studies focused on languages other than English [7], [8].

We do not have a comprehensive understanding of the ‘normal’ developmental trajectory of lexical stress contrasts during speech production. Perhaps this can be said of a number of other aspects of speech production [9]. An understanding of the normal or typical trajectory of development is essential for understanding atypical or delayed development.

The normalized pairwise variability index (PVI) formula allows for standardized comparisons of stress contrastivity in terms of duration, intensity, and fundamental frequency. Even when children’s word productions are judged by an adult listener as being intelligible and correct, there could be fine-grained acoustic differences in the magnitude of stress contrastivity that differentiate child and adult speech.

It is important to note that our focus here is on what acoustic analyses can reveal about changes in how lexical stress contrasts are realized during child development. Theories of why some syllables attract stress in English are not aligned with the research we report here. Speech-motor theories of how speech is produced (including segmental and suprasegmental aspects of speech) are more aligned with the research we report here. However, very few, if any, of these theories consider development (much less the magnitude of stress contrasts and why physiological and/or practice effects may affect children’s capacity to produce adult like magnitudes of contrasts). As such we are important for building developmentally relevant theory.

The so called trochaic bias suggests that words beginning with a weak syllable may present more challenges than words beginning with a strong syllable. Based on our previous research in English using the PVI to measure lexical stress contrasts in stimuli elicited via picture naming, but with a narrower age range and fewer stimuli, we anticipated that we might observe differences in children versus adults in WS word productions [10], [11]. However we note that much of the trochaic bias research has come from research on English and there has been debate about the universality of this bias, a point we touch on again briefly in the Discussion.

2. Method

A total of 261 children aged between 3 and 12 years of age were recruited. Of these 8 were excluded leaving 253 children (mean age = 94.66 months, SD = 32.50, range = 36-150), 128 girls and 125 boys. Adults included 23 eligible participants (mean age = 19.7 years, SD = 3.3; range 18-29), 22 females and 1 male. Sample size for each age as follows. For 3 years of age: 15 participants; 4 years: 32; 5 years: 27; 6 years: 24; 7 years: 29; 8 years: 24; 9 years: 29; 10 years: 26; 11 years: 24; 12 years: 20; adult: 23.

Participants wore a headset microphone at a 10cm distance from the mouth. Speech was recorded to a handheld Zoom H4N Handy Recorder digital recorder (44 kHz sampling rate, 16 bit quantization). A picture naming task was used to elicit word productions 12 with WS stress (banana, bandana, bikini, cathedral, computer, confetti, potato, pyjamas, spaghetti, tomato, tornado, zucchini). These words were elicited twice (i.e., lists with different word orders were used so that two productions were elicited per word over the course of the recording session). Note that in Australian English, this task elicits a single word utterance the vast majority of the time. In some languages, the task may elicit an article as well as the target but this is not the case in our study.

Word frequency values for all words were identified using the Sublex-UK word frequency database (van Heuven et al., 2014). Frequency values are reported in the form of a logarithmic scale, hereafter referred to as the Zipf scale, with values between 1-3 indicating low frequency while 4-7 indicate high frequency. Words used for this study were found to be in the mid frequency range with a mean of 3.67 for words with WS stress [17]. Note this was part of a larger study.
that also included strong-weak (SW) words. Only the WS are reported in this conference paper. Only correct responses for the WS (and SW) stimuli were analysed acoustically. Usable recordings were captured for 14,315 of a possible 14,904 productions (96% of the data). Those recordings deemed to be unusable contained productions which were either too loud or too quiet, or were obscured by excessive background noise (e.g., movement of the microphone). Each word production was judged correct or incorrect based on phonemic composition and stress pattern by a trained research assistant. Of the usable recordings, a total of 12,914 individual word productions (90%) were correct or near-correct (e.g., elongation [e.g., pyjamas with prolonged “z”]). The WS data are a subset of this larger dataset.

Semi-automated segmentation procedures were then undertaken using the Munich Automatic Segmentation tool (MAUS) as described by Arciuli et al. [12]. The recording of each correct word production obtained from our participants was uploaded to the webMAUS interface (version 3.11), resulting in a TEXTGRID file. These phonemic segments were then reviewed and adjusted manually by a trained research assistant using waveforms and wide-band spectrograms with a 300-Hz bandwidth.

For each correct production, the resulting TEXTGRID and .wav file pairs were analysed using a script created in the Praat software. This script generated duration (ms), and mean and maximum pitch (Hz) and intensity (dB) for each vowel defined by the TEXTGRID file. Duration values were defined as the difference between the onset and offset of the vowel. Pitch values were defined at a minimum of 75 Hz and maximum of 900 Hz and intensity values were defined at 50 dB to 100dB. Otherwise, the standard settings in Praat were used. The parabolic interpolation algorithm in Praat was used to estimate the location and value of the maximum pitch. Mean intensity values reflected the mean of the intensity curve, while maximum intensity values were calculated according to the cubic interpolation method in Praat. This process and reliability data is reported in more detail by Arciuli and colleagues [12].

These vowel statistics were used to calculate a normalized Pairwise Variability Index (PVI) [13] the magnitude of contrast in the vowel statistics of the first two syllables within each word production (i.e., stress contrastivity). While some other studies have reported vowel statistics on their own we were specifically interested in the magnitude of contrast. The PVI value represents a normalised difference between the first two vowels of each word and was calculated using the formula below (Equation 1), where $a_1$ and $a_2$ represent duration, peak intensity, or peak $f_0$ for the first and second vowel, respectively:

$$PVI = 100 \times \left( \frac{a_1 - a_2}{(a_1 + a_2)/2} \right)$$

A negative PVI value represents a WS stress pattern, while a positive PVI value represents a SW stress pattern. Mean normalised PVI values were computed to create three grand PVI values for each participant: (i) PVI_duration_WS, (ii) PVI_intensity_WS, and (iii) PVI_ $f_0$ WS.

As per Ballard et al. [10] and Arciuli and Ballard [11], univariate analyses of variance (ANOVPAs) were conducted. Effect sizes ($η^2$) were small (.01), medium (.06), or large (.14) [14]. Note: as the assumption of homogeneity of variance was not met, Welch’s ANOVAs were used. We used reverse-Helmert contrasts to compare each level of ‘age’ to the pooled mean of previous levels (e.g., adults vs. all children, 12-year-olds vs. all younger children, etc.). Due to the number of comparisons, we used an adjusted alpha of .01.

### 3. Results

Raw data for each correct word production were examined and outliers more than three standard deviations from mean scores, generated for the six grand PVI values using the 12,914 word productions, were identified and removed. For WS words a total of 36 PVI duration, 156 PVI intensity, and 54 PVI $f_0$ values were removed (3.9% of data). We averaged across items, for participants in each age group, to derive descriptive statistics. As expected, PVIs for WS words were generally negative (see Table 1).

#### Table 1: Mean (SD) PVIs by age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>PVI_duration (SD)</th>
<th>PVI_intensity (SD)</th>
<th>PVI_ $f_0$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 years</td>
<td>97.18 (44.70)</td>
<td>1.82 (10.30)</td>
<td>-11 (32.17)</td>
</tr>
<tr>
<td>4 years</td>
<td>-93.54 (52.25)</td>
<td>-4.72 (12.12)</td>
<td>-2.40 (32.86)</td>
</tr>
<tr>
<td>5 years</td>
<td>-91.85 (52.07)</td>
<td>-3.36 (11.34)</td>
<td>-3.22 (35.32)</td>
</tr>
<tr>
<td>6 years</td>
<td>-96.04 (48.42)</td>
<td>-2.65 (12.08)</td>
<td>-3.03 (31.66)</td>
</tr>
<tr>
<td>7 years</td>
<td>-94.47 (48.09)</td>
<td>-4.20 (13.53)</td>
<td>-2.27 (38.63)</td>
</tr>
<tr>
<td>8 years</td>
<td>-91.52 (49.34)</td>
<td>-5.51 (14.23)</td>
<td>-3.95 (39.67)</td>
</tr>
<tr>
<td>9 years</td>
<td>-94.57 (47.36)</td>
<td>-5.41 (14.97)</td>
<td>-7.56 (42.83)</td>
</tr>
<tr>
<td>10 years</td>
<td>-89.80 (44.98)</td>
<td>-4.44 (13.20)</td>
<td>-1.28 (38.44)</td>
</tr>
<tr>
<td>11 years</td>
<td>-93.40 (45.82)</td>
<td>-6.18 (15.41)</td>
<td>-3.61 (40.14)</td>
</tr>
<tr>
<td>12 years</td>
<td>-91.44 (45.59)</td>
<td>-5.10 (11.39)</td>
<td>2.22 (31.29)</td>
</tr>
<tr>
<td>Adult</td>
<td>-98.77 (48.24)</td>
<td>0.09 (2.98)</td>
<td>4.87 (7.29)</td>
</tr>
</tbody>
</table>

Correlational analyses revealed statistically significant relationships between Zipf frequency and PVI duration ($p < .01$) as well as PVI intensity ($p < .01$) for WS word productions. In these cases Zipf frequency was included in additional covariate analyses.

There was no main effect of age for PVI duration scores calculated using WS word productions ($F[10,180] = .05, p = 1.0, η^2 = .003$). A follow-up ANCOVA was conducted with Zipf frequency as a covariate. It did not reveal a statistically significant effect of age, $F[10,179] = .08, p = 1.0, η^2 = .004$). In some of our previous studies we restricted WS words to those beginning with a schwa. Removal of the two items that do not begin with a schwa did not affect the pattern of results.

For PVI intensity, the main effect of age group was statistically significant ($F[10,180] = 4.28, p < .001, η^2 = .19$). A follow-up ANCOVA was conducted with Zipf frequency as
a covariate and it also revealed a statistically significant effect of age, \(F_{[10,179]} = 4.70, p < .001, \eta^2 = .21\). The reverse Helmert contrasts showed a statistically significant difference between the adults versus the entire group of children. These effects of age group did not change substantially when we removed the two words that do not begin with a schwa and re-ran the analyses.

4. Discussion

While many of the children’s productions were intelligible, and the mean PVIs were appropriately negative for WS words (indicating greater stress on the second syllable relative to the first), we found considerable variability in the amount of stress contrastivity produced by children. The most consistent finding was that children had not yet achieved adult-like stress contrastivity for our acoustic measures of intensity and fundamental frequency for WS words. In terms of duration, children generally showed similar stress contrastivity to adults. However, when children were viewed as a group and compared with adults they generally showed greater stress contrastivity than adults in terms of intensity and fundamental frequency. We reported similar findings for WS words in previous studies but those studies included a narrower age range and fewer stimuli [10, 11].

Perhaps there are physiological constraints in the maturing speech production system. Redford [9] pointed out that early onset of speech combined with some late maturing mechanisms represents something of a paradox. Kent and Forner [15] noted variability in the duration of children’s segmental productions up to 12 years of age. Smith and Zelaznik [16] reported that control of articularatory timing continues into adolescence. Redford’s review and the studies mentioned above were not focused on stress contrastivity but did indicate that some aspects of speech production may be late developing. Another possibility is that different kinds of physiological effects might be at play for younger versus older children, but that the result is the same in that younger and older children are still not adult-like in their production of lexical stress.

Perhaps there are practice effects at play here as well given the WS pattern of lexical stress is less dominant in English. The issue of the so called trochaic bias and the interplay between physiological and practice effects has been explored in a study of lexical contrastivity in typically developing children who speak Italian as words beginning with a weak syllable are more common in Italian than English [7]. Note that the possibility of this kind of practice effect is different from the issue of the frequency of individual items.

We examined the frequency values of individual items and, where appropriate, took them into account via covariate analyses. The results suggest that it is highly unlikely that the individual frequency of particular items can explain any of the effects we report here.

Clearly, there is more work to be done on understanding why lexical stress contrastivity, at least for the WS stimuli we report here, is late maturing. We recommend that an increased number of cross-linguistic studies be conducted so that we might gain a more comprehensive understanding of the development of lexical stress contrastivity during speech production. There have been some cross-linguistic studies in this area but more research effort is needed.

5. Conclusions

Our results suggest a protracted developmental trajectory for production of lexical stress contrastivity. Even when typically developing children are producing intelligible and correct word productions which exhibit stress contrastivity and they are not necessarily producing adult-like magnitudes of contrastivity in their production of WS words. This is the largest acoustic study we know of to examine the developmental trajectory of this aspect of speech production and will assist in building developmentally relevant theories of lexical stress and of speech production. Similar acoustic data for production of SW words in English, and an increase in cross-linguistic research, is required before we can speculate further about the reasons for this protracted developmental trajectory.

6. Acknowledgements

The research reported here is part of a larger project funded by a grant from the Australian Research Council awarded to Arciuli, Ballard, and Vogel.

7. References


