Using MAUS to Investigate Children's Production of Lexical Stress

Joanne Arciuli[1], Kirrie Ballard[1], Katelyn Phillips[1,2], & Adam Vogel[3]

[1]The University of Sydney, Australia
[2]Monash University, Australia
[3]The University of Melbourne, Australia
Joanne.arciuli@sydney.edu.au

ABSTRACT

Contrasting strong versus weak syllables within words (lexical stress) is critical for effective communication in English. Yet acoustic data required to shed light on stress contrastivity are usually obtained via laborious manual methods, an obstacle to large-scale studies. The automatic alignment procedure in the Munich Automatic Segmentation tool (MAUS) [1] might reduce the manual effort required for acoustic analyses. However, there is little data on the reliability of MAUS when compared with manual measurements in analysing child speech. We report on a subsample taken from a large study designed to investigate lexical stress production in typically developing Australian English-speaking children. We compared manual acoustic measurements with measurements obtained via MAUS. The results from analysis of 200 word productions showed moderate to high correlations between the measurements. However, MAUS tended to overestimate the duration of weak vowels (but not strong vowels). Use of MAUS in combination with manual checks is recommended.

Keywords: lexical stress, prosody, acoustic analysis, MAUS, children's speech.

1. INTRODUCTION

All languages exhibit rhythm or prosody. As part of its prosodic system, English utilizes lexical stress: the contrast between strong and weak syllables within words (compare the strong-weak pattern of lexical stress in 'INcense' with the weak-strong pattern of 'inCENSE'). In languages such as English, lexical stress is important for intelligibility. Despite knowing of its importance for effective communication, we know relatively little about how stress contrastivity might change with development.

Of the small amount of previous research that has examined lexical stress production in typically developing children much of it reports on syllable truncation/preservation, and stress shift, rather than stress contrastivity per se [2-4]. We know that the production of stress contrastivity begins in infancy during babbling [5] and that by 3 years of age typically developing children use duration, intensity and fundamental frequency to contrast syllables within words; although, they have not necessarily reached adult-like intentional control of lexical stress contrastivity [6-8]. In fact, recent acoustic studies suggest that there is a far more protracted developmental trajectory for adult-like mastery of lexical stress production than previously thought [9, 10].

Large acoustic investigations of stress contrastivity are extremely rare because of the technical and laborious nature of manual acoustic analyses. Here we report on a large study designed to investigate lexical stress production in typically developing Australian English-speaking children. The data we report compares manual acoustic measurements with those obtained via an automatic alignment procedure obtained via the Munich Automatic Segmentation tool (MAUS) [1].

A search of the databases CINAHL, Scopus, Medline and PsychInfo was conducted in order to identify studies using MAUS for the analysis of children's speech. The search terms included: 'MAUS', 'automatic segmentation', 'phonemic segmentation', 'automatic alignment' AND 'speech'. Articles including the keyword 'reading' were excluded due to the overlap with articles concerning the development of phonemic segmentation for reading. If an article appeared relevant from the title, the abstract was read to establish relevance. These articles were scanned for the search terms 'MAUS' or 'Munich'. In addition to this database search, any publications that cited any publications by the authors of MAUS were searched to find articles that used MAUS to analyse children's speech.

From all relevant sources, only two publications were found that used MAUS software in child-related research; those by Falk and colleagues [11] and Peters [12]. Both articles used the MAUS software to segment audio recordings for children and adolescents, with subsequent manual correction of any errors in segment boundaries. Falk and colleagues [11] presented eight adolescents aged 11 to 15 who stutter with two tasks, comparing sung and spoken utterances in German. The results of this indicated that Voice Onset Times are reduced in sung compared to spoken utterances. Peters [12] presented a picture naming task to eight typically developing children aged 5 to 7 years. The target utterances were 12 simplex or complex Standard Southern German words. Neither study compared manual acoustic measurements to MAUS analysis.

This literature review highlights an absence of studies comparing manual acoustic measurements to measures obtained using MAUS in the analysis of children's speech production. As far as we are aware, no previous study has used MAUS to examine suprasegmental aspects of children's speech production such as lexical stress contrastivity. Our aim here is to provide data that demonstrates the reliability of MAUS for this kind of acoustic investigation of children's speech.

2. METHOD

2.1. Participants

Children aged between 3 and 12 years of age were recruited from ten daycare centres, preschools and primary schools in Sydney, Australia. Children were included in the study if they attended English speaking schools and had spoken English for longer than a year. They were excluded if they had a history of a speech or communication disorder, including language disorder and autism spectrum disorder, according to parental report.

The current study reports data from a subset of participants from a larger sample recruited for an investigation of prosody during speech production. A random sample of 200 correct word productions from 166 children with a mean age of 100.6 months (SD = 29.8, range = 36-148 months) were used for the current study; 85 girls (and 81 boys).

The study was approved by the Human Research Ethics Committee of the University of Sydney and, where appropriate, the New South Wales Department of Education and Communities (for public schools), or the Catholic Education Office (for Catholic schools). All center managers, school principals, and parents provided written consent for children to participate in this study.

2.2. Experimental task

Children were tested individually in a quiet room on school premises. A picture naming task was used to elicit word responses. If naming did not occur on presentation of the picture stimulus, responses were prompted by the examiner – first a description of the picture was given, followed by a phoneme cue. If naming still did not occur, a spoken model of the target word was provided. Order of presentation of stimuli was varied using four different word lists. Each child completed the picture naming task twice using two different lists. Children wore a headset microphone at a 10cm distance and speech was recorded to a handheld Zoom H4N Handy Recorder digital recorder (44 kHz sampling rate, 16 bit quantization).

2.3. Speech stimuli

Target words included both strong-weak (SW) and weak-strong (WS) stress patterns across the initial two syllables. For each word, the first two syllables were included in the analysis. Polysyllabic rather than bisyllabic words were chosen to avoid measurement of syllables where there may be word-final lengthening. There were 27 words in total; 15 SW words and 12 WS words. Targets are listed in Table 1.

To facilitate acoustic analysis, stimuli were selected with the following constraints: (1) all followed the same phonological structure in that the first two vowels being measured were embedded between consonants, (2) all contained consonants that have been found to be in the consonant inventory of young children and (3) all had easily demarcated vowel onsets and offsets in the acoustic signal (e.g., no liquids or semivowels). All stimuli were names of picturable objects. Adhering to these constraints made it more difficult to select WS words which is why there are slightly fewer of those kinds of words.

Table 1: Target words for study. Note: *vegemite is an Australian food spread.

Strong-Weak Targets	Weak-Strong Targets
barbecue	banana
bicycle	bandana
butterfly	bikini
caterpillar	cathedral
coconut	computer
cucumber	confetti
dinosaur	potato
hamburger	pyjamas
motorbike	spaghetti
newspaper	tomato
photograph	tornado
pineapple	zucchini
porcupine	
saxophone	
vegemite*	

2.4. Acoustic measurements

Recordings of each word production were segmented into individual word productions and saved in .wav format using PRAAT software [Version 5.3.78; 13]. Each of these word productions was listened to and judged correct based on phonemes by a trained research assistant. Following this, a random sample of 200 correct word productions (100 SW and 100 WS) was chosen from the overall sample. These words were each analysed twice, once using MAUS, and once using manual measurements obtained via PRAAT software [13] alone. The manual measurements were conducted by a trained research assistant.

As noted, the MAUS system is a tool designed to automatically segment speech according to phonemes [1]. Given a speech signal and a related orthographic representation, MAUS estimates the most likely pronunciations from canonical pronunciation and finds the most likely phonemic segmentation. Originally developed in German, MAUS has been adapted into several languages including English. The recording of each correct word production obtained from our participants was uploaded to the webMAUS interface (version 3.11) [14], resulting in a TEXTGRID file. These TEXTGRID files were used to generate durations for vowel 1 (V1) and vowel 2 (V2) for each word in PRAAT.

For manual measurements, waveforms and wide-band spectrograms with a 300-Hz bandwidth were generated for each sound file in PRAAT. Vowel duration was measured from the onset to offset for V1 and V2 [15].

These duration measurements were then used to calculate a normalized Pairwise Variability Index (PVI) [16] reflecting the amount of contrast in the vowel durations of the first two syllables within each word production (i.e., stress contrastivity). 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 a1 and a2 represent duration for the first and second vowel, respectively:

$$PVI\_a = 100 \times \{(a_1 - a_2) \div [(a_1 + a_2) \div 2]\} \quad (1)$$

The normalized PVI formula allows for standardized comparisons between participants. It has been used to examine lexical stress contrastivity in typically developing children's speech production [9, 10, 17], in a study of children's apraxia of speech [18], and a recent study of speech production in children with and without autism spectrum disorders [19]. A negative PVI value represents a WS stress pattern, while a positive PVI value represents a SW stress pattern.

2.5. Statistical analysis

All statistical analyses were conducted using SPSS. Analyses were conducted separately for SW and WS vowels.

Parametric correlations using Pearson's r were conducted to examine the relationship between measurements obtained manually versus those obtained using MAUS, for both V1 and V2. An alpha level of .01 was used for the correlation analyses.

3. RESULTS

The 200 correct word productions (100 SW and 100 WS) chosen for this analysis were randomly selected from 166 children that came from a larger pool of participants. Due to random sampling, some participants provided multiple words amongst the 200 selected.

The average durations for each vowel and word type as measured both manually and by MAUS are represented in Table 2.

Table 2: Correlations between vowel durations by analysis type. Note: * indicates significant at the 0.01 level.

Word/ Vowel type	Manual Mean (SD)	MAUS Mean (SD)	Correlation (r)
SW vowel 1	94.8 (*40.0*)	94.8 (*53.6*)	.82*
SW vowel 2	58.7 (*38.9*)	68.9 (*42.8*)	.88*
WS vowel 1	48.4 (*33.1*)	53.9 (*35.5*)	.71*
WS vowel 2	135.3 (*49.1*)	137.5 (*60.4*)	.80*

Measurements of vowel duration derived manually were statistically significantly correlated with measures obtained by MAUS. Correlation coefficients were lowest for the first vowel of WS vowels words ($r = .71$, $p < 0.01$) and highest for the second vowel of SW ($r = .88$, $p < 0.01$). Scatterplots for these correlations are shown in Figures 1 and 2.

Figure 1: Scatterplots of correlations between vowel durations as calculated by hand and by MAUS for Strong-Weak words.

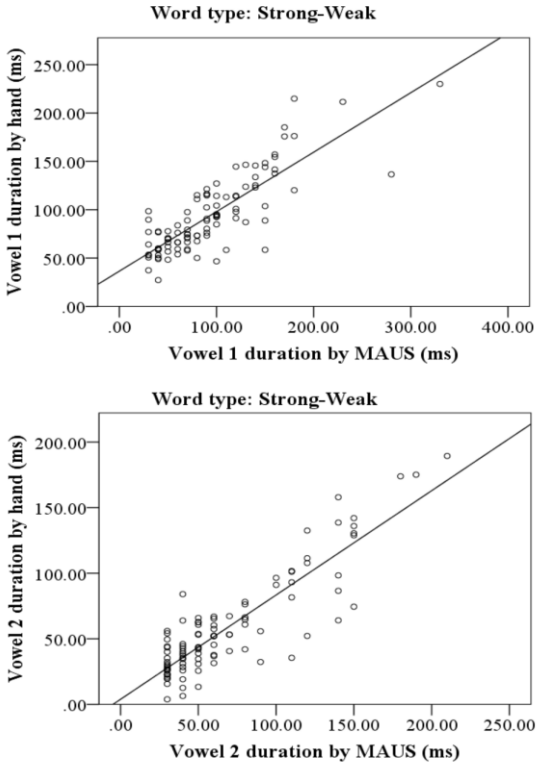
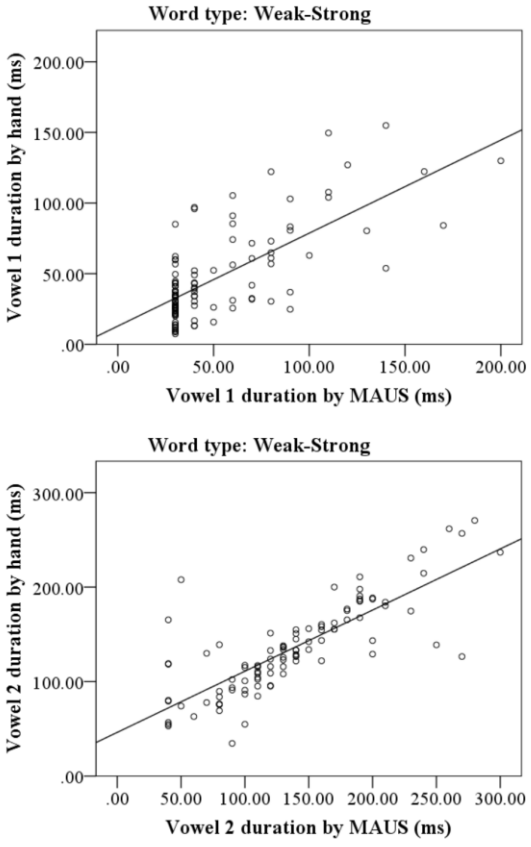


Figure 2: Scatterplots of correlations between vowel durations as calculated by hand and by MAUS for Weak-Strong words.



Paired samples t-tests were also performed on the vowel durations as calculated by hand and by MAUS. According to this analysis, the vowel durations calculated by MAUS were on average 10.21 milliseconds longer than durations calculated by hand for the second vowel of SW words (the weak vowel). This difference was significant ($t(99) = -4.92$, $p < .01$). For WS vowels, the vowel durations calculated by MAUS were on average 5.53 milliseconds longer than durations calculated by hand for the first vowel (the weak vowel). This difference was significant ($t(99) = -2.10$, $p = .04$). No other differences were statistically significant. These results are summarized in Table 3.

Table 3: Difference in vowel durations obtained by manual versus MAUS measurements. Note: * indicates significant at the 0.05 level.

Word/ Vowel type	Mean Difference (SD)
SW vowel 1	.01 (*30.65*)
SW vowel 2	10.21 (*20.73*)*
WS vowel 1	5.53 (*26.34*)*
WS vowel 2	2.21 (*36.50*)

The results of the correlational analysis conducted to examine the relationship between PVI values derived from manual durations versus those obtained using MAUS can be found in Table 4. There was a significant correlation between manual PVIs and MAUS PVIs for both SW ($r = .74$, $p < 0.01$) and WS words ($r = .70$, $p < 0.01$).

Table 4: Correlations between PVI values by analysis type. Note: * indicates significant at the 0.01 level.

Word type	Manual PVI Mean (*SD*)	MAUS PVI Mean (*SD*)	Correlation (r)
SW	53.1 (*59.2*)	31.3 (*66.4*)	.74*
WS	-95.5 (*58.4*)	-83.0 (*63.1*)	.70*

4. DISCUSSION

Here we compared manual acoustic measurements with those obtained via MAUS for 200 correct word productions from typically developing children. Results showed moderate to high correlations between vowel duration data obtained from these methods. However, we also observed a tendency for MAUS to overestimate the duration of weak vowels in children's speech. In view of this, we suggest that MAUS is useful for acoustic measurements of segmental durations and features although some manual corrections might be advisable regarding the boundaries of weak vowels.

The current study reports on a subset of data taken from a large study of Australian English speaking children designed to investigate the normal developmental trajectory of lexical stress production. There is growing interest in this area as recent research has shown a more protracted trajectory for typically developing children to reach adult-like mastery of stress contrastivity than previously thought [9, 10]. Examining stress contrastivity is challenging due to the laborious nature of acoustic analyses that examine the key suprasegmental features of lexical stress (i.e., vowel duration, intensity, and fundamental frequency – all of which require the identification of phoneme boundaries in the speech stream as an initial step). Yet, acoustic analysis is the only data that can reveal fine-grained developmental changes in stress contrastivity (e.g., changes relating to the degree of contrast in vowel durations across strong versus weak syllables with single words).

The MAUS procedure can reduce the time required for acoustic analyses by providing automated identification of phoneme boundaries in the speech stream. However, there is little data on how acoustic data obtained via MAUS compares with that obtained via manual methods. Our search revealed only two previous studies that have used MAUS to examine children's speech production [11, 12]. Neither of these studies compared MAUS with manual methods and neither examined acoustic features relating to stress contrastivity.

In the current study we observed a statistically significant tendency for MAUS to overestimate the duration of weak vowels, which are very brief events often with low intensity, suggesting that some manual corrections are needed for these vowels before reliable values can be calculated when analysing children's speech.

In addition, as far as we are aware, MAUS segments vowels to only two decimal places, meaning that resulting PVI values can occasionally be zero (first and second vowels are identified as being the same length to two decimal places), unlike the PVI values that result from manual measurements that reflect more decimal places. When segmentation is done by hand, PVI values are very rarely zero. This could account for lower correlations for PVI values reported in Table 4 by comparison with some of the individual vowel measurements reported in Table 2.

5. ACKNOWLEDGEMENTS

This research was funded by a Discovery Project from the Australian Research Council awarded to Joanne Arciuli, Kirrie Ballard and Adam Vogel (DP130101900). We thank the research assistants, schools/centres, parents, and children who took part in this research.

6. REFERENCES

[1] Schiel, F. 1999. Automatic phonetic transcription of non-prompted speech. *Proc. 14th ICPhS* San Francisco, California, 607-610.

[2] Jarmulowicz, L. D. 2002. English derivational suffix frequency and children's stress judgments. *Brain and Language* 81, 192-204.

[3] Kehoe, M. M. 2001. Prosodic patterns in children's multisyllabic word productions. *Language, Speech, and Hearing Services in Schools* 32, 284-294.

[4] Roy, P., Chiat, S. 2004. A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research* 47, 223-234.

[5] Davis, B.L., MacNeilage, P. F., Matyear, C. L., Powell, J. K. 2000. Prosodic correlates of stress in babbling: An acoustical study. *Child Development* 71, 1258-1270.

[6] Kehoe, M., Stoel-Gammon, C., Buder, E. H. 1995. Acoustic correlates of stress in young children's speech. *Journal of Speech, Language, and Hearing Research* 38, 338-350.
[7] Pollock, K. E., Brammer, D. M., Hageman, C. F. 1989. An acoustic analysis of young children's production of word stress. *Journal of Phonetics* 21, 183-203.
[8] Schwartz, R. G., Petinou, K., Goffman, L., Lazowski, G., Cartusciello, C. 1996. Young children's production of syllable stress: An acoustic analysis. *The Journal of the Acoustical Society of America* 99, 3192-3200.
[9] Arciuli, J., Ballard, K. J. 2017. Still not adult-like: Lexical stress contrasivity in word productions of eight- to eleven-year olds. *Journal of Child Language* 44, 1274-1288.
[10] Ballard, K. J., Djaja, D., Arciuli**,** J., James, D., van Doorn, J. 2012. Developmental trajectory for production of prosody: Lexical stress contrastivity in children 3 to 7 years and adults. *Journal of Speech, Language, and Hearing Research* 55, 1822-1835.
[11] Falk, S., Maslow, E., Thum, G., Hoole, P. 2016. Temporal variability in sung productions of adolescents who stutter. *Journal of Communication Disorders* 62, 101-114.
[12] Peters, S. M. 2015. The effects of syllable structure on consonantal timing and vowel compression in child and adult speakers of German. Ph.D. dissertation, Ludwig–Maximilians–Universitat Munchen, Munich, Germany.
[13] Boersma, P., Weenink, D. 2017. Praat: Doing phonetics by computer. [Computer program]. Version 5.3.78, retrieved 11 February 2017.
[14] Kisler, T., Schiel, F., Sloetjes, H. 2012. Signal processing via web services: The use of case WebMAUS. *Proc. Digital Humanities Conference* Hamburg, Germany.
[15] Peterson, G. E., Lehiste, I. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32, 693-703.
[16] Nolan, F., Asu, E. L. 2009. The pairwise variability index and coexisting rhythms in language. *Phonetica* 66, 64-77.
[17] Arciuli, J., Colombo, L. 2016. An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. *Speech Communication* 80, 22-33.
[18] Ballard, K. J. Robin, D. A., McCabe, P. McDonald, J. 2010. A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research* 53, 1227-1245.
[19] Arciuli, J., Bailey, B. 2018. An acoustic study of lexical stress contrastivity in children with and without autism spectrum disorders. *Journal of Child Language*. Advance online publication.