Xu Rattanasone, N., Yuen, I., Holt, R., & Demuth, K. (2022). Jellybeans… or Jelly, Beans…? 5-6-year-olds can identify the prosody of compounds but not lists. *Journal of Child Language, 49*(3), 602-614.

Access to the published version:
https://doi.org/10.1017/S0305000921000234

Running Head: ACQUISITION OF COMPOUND PROSODY

***Jellybeans*… or *Jelly, Beans*…? 5-6-year-olds can identify the prosody of compounds but not lists**

Nan Xu Rattanasone, Ivan Yuen, Rebecca Holt, and Katherine Demuth

Department of Linguistics and Centre for Language Sciences, Macquarie University

Corresponding author: Nan Xu Rattanasone (email: nan.xu@mq.edu.au)

Acknowledgements

We thank Gretel MacDonald, Stefanie Shattuck-Hufnagel and members of the Child Language Lab at Macquarie University for helpful comments and suggestions, and Peter Petocz for statistical advice. Partial funding for this research was provided by the following grants: ARC FL 130100014 (Demuth) and ARC Centre of Excellence for Cognition and its Disorders CE110101021.

Abstract

Learning to use word versus phrase level prosody to identify compounds from lists is thought to be a protracted process, only acquired by 11 years (Vogel & Raimy, 2002). However, recent study has shown that 5-year-olds can use prosodic cues other than stress for these two structures in *production*, at least for early acquired noun-noun compounds (Yuen et al., 2021). This raises the question of whether children this age can also use naturally-produced prosody to *identify* noun-noun compounds from their list forms in comprehension. The results show that 5-6-year-olds (N=28) can only identify compounds. Unlike adults, children as a group could not use boundary cues to identify lists and were significantly slower in their processing compared to adults. This suggests that the acquisition of word level prosody may precede the acquisition of phrase level prosody, i.e., some higher-level aspects of phrasal prosody may take longer to acquire.

(147 words)

Jellybeans… **or** ***Jelly, Beans…?*** **5-6-year-olds can identify the prosody of compounds but not lists**

In English, a language where compounding is highly productive, children have been reported to construct novel compounds, e.g., *nose-beards* for moustaches, and exhibit compound stress in their production at around 2 to 3 years (Becker, 1994; Clark, 1981; Clark, Gelman, & Lane, 1985). Although 3-5-year-olds show good semantic understanding of familiar compounds, e.g., that *chocolate-cake* is a cake made of chocolate (Krott & Nicoladis, 2005; Nicoladis, 2003), 5-year-olds cannot identify *unfamiliar* compounds like *wetscrew* in comprehension (Vogel & Raimy, 2002). Even in studies using *familiar* compounds, e.g., *chocolate-cake*, there is still substantial individual variation among children's performance, ranging from 43.8% to 93.8% correct (Wells, Peppé, & Goulandris, 2004). This variable performance across different studies may be attributable to either the semantic understanding of different types of compounds, or children's emerging ability to use prosodic cues (e.g. stress vs. duration, pauses) in identifying compounds. This study re-examines children's use of prosodic cues in the comprehension of early acquired noun-noun compounds in an online eye-tracking task to further probe their knowledge of the prosodic structure of compounds.

Compounds are prosodically complex, composed of two prosodic words (PW) to form a new prosodic word: [[jelly]$_{PW}$ [beans]$_{PW}$]$_{PW}$ (cf. Wheeldon & Lahiri, 2002; Wynne, Wheeldon, & Lahiri, 2018). Treating a compound as a single prosodic word will help children correctly identify '*jellybeans* and *chips*' as a two-item list. However, treating a compound as two separate prosodic words will lead children to incorrectly identify it as a three-item list '*jelly*, *beans* and *chips*'.

In English, compounds and their list forms differ in word (i.e. lexical) stress and boundary-marking, which manifest through respective acoustic cues (e.g., pitch and duration) in production. According to the compound stress rule, the second word of a compound loses its primary stress, resulting in a strong-weak pattern for '*JEllybeans*', while retaining it in the list form, resulting in a strong-strong pattern for '*JElly, BEANS*' (Chomsky & Halle, 1968). Since primary stress is often acoustically realised with high pitch, the second word in a compound will be lower in pitch than the stressed counterpart in list form, i.e., *beans* in "*jellybeans*" will be lower in pitch than in "*jelly*, *beans*".

By two-and-a-half (2;6) years, English-speaking children correctly assign primary stress to the initial noun of noun-noun (N-N) compounds, and 2- to 4-year-olds can produce the appropriate strong-weak stress pattern for 95% of novel N-N compounds (Clark et al. 1985). Similarly, by 2;6 years, children can correctly interpret the meaning of novel N-N compounds, e.g., by associating *apple-knife* with a *knife* and not an *apple* (Clark et al. 1985). This shows that young children can use stress information to derive the meaning of N-N compounds, i.e., that *knife* is the head of the compound *apple-knife*, and *apple* is the modifier.

Apart from word stress, boundary cues can also help identify compounds. When a compound is produced as a single prosodic word, there should not be any word-internal prosodic boundary (Cutler & Butterfield, 1990). This means pauses are less likely to be found in compounds. In contrast, for lists, each item is likely to be treated as a phrase and marked with a pause and boundary. The presence of a phrase boundary is typically also accompanied by pre-boundary lengthening, realised as longer duration in the final syllable of a word before a phrase boundary (Snow, 1994; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Therefore, the second syllable of "*jelly*" will only be lengthened in the three-item list form but not as a compound, e.g., ***jelly***, ***beans*** & *chips* vs ***jellybeans*** *&* *chips* (where underlined items should have pre-boundary lengthening).

A recent production study showed that 5-year-olds can *produce* different prosodic cues (pitch, duration) N-N compounds and their list forms (e.g., *jellybeans* vs. *jelly*, *beans*) similar to adults, suggesting an emerging representation for word and phrase level prosodic structures (Yuen, et al., 2021). However, children were more likely than adults to use pauses within compounds, suggesting that their representations for boundary cues may not yet be robust.

Children's ability to perceive boundary cues is inconsistently reported in the literature. One study reported that 3-day-old infants can discriminate between within- vs. between-word disyllables (Christophe, Dupoux, Bertoncini, & Mehler, 1994). However, much older children reportedly cannot use boundary cues to identify a novel compound vs. two separate prosodic words (in a noun phrase), e.g., *wetscrew* vs. *wet screw*, in a two-alternative forced-choice task (Vogel & Raimy, 2002). Five-year-olds could not identify compounds either, suggesting a bias against compound interpretation (Vogel & Raimy, 2002). This finding seems inconsistent with the study of much younger children showing that even infants are sensitive to lexical stress and boundary cues in comprehension. One explanation might be that transparency in meaning might influence how prosodic cues are evaluated in a task that requires an explicit response from children. Thus, while young children show good semantic understanding of N-N compounds, they show a lack of understanding for other types of compounds in early development. For example, ill-formed compounds in early production often involve verbs (V-N) e.g., *cracking-nut* (2;6) for 'nut-cracker' (Clark et al., 1986). Even at 6;5, children are not productive in generating novel compounds involving inanimates (Clark et al., 1986), the compound type that Vogel and Raimy (2002) used. This suggests that compounds other than the N-N variety may be semantically more challenging for young children. Perhaps they can perceive and produce the correct prosodic cues for earlier acquired N-N compounds before they can do so for other compound types. While compound stress cues occur at the lexical level, durational cues occur at the phrase level. If the inconsistent reports of cue use in previous research is due to children's (in)ability to use prosodic cues at different levels of structure, then young children should show challenges even in identifying earlier acquired N-N compounds.

The Current Study

The current study therefore examined whether 5-6-year-olds can identify N-N compounds from their list forms while listening to naturally produced speech. We used an Intermodal Preferential Looking paradigm to explore the online processing of compounds vs. their list forms, an area of research which is relatively underexplored. This paradigm has the advantage of not requiring participants to make overt response, while providing information on how prosodic cues are processed over time. One study using eye-tracking with adult Japanese listeners exploited the Compound Accent Rule (CAR) to see if adults could use the rule to predict the presence of compounds vs. non-compound forms (Hirose & Mazuka, 2015). In Japanese, when CAR is applied, many words undergo changes in pitch, e.g., words starting with High pitch are reassigned Low pitch when they are produced as part of a compound (Hirose & Mazuka, 2015). For those words, adult listeners can unambiguously determine when a compound word is expected after hearing just the first word of the compound (Hirose & Mazuka, 2015). This suggests that listeners are actively monitoring pitch information to help predict upcoming words. Another online study using electroencephalography (EEG) with English-speaking adults found similar results, with misapplication of lexical level (compound) prosody eliciting an earlier (N400) response than inappropriate phrase level (list) prosody, which elicited a P600 (McCauley, Hestvik, & Vogel, 2013). These results suggest that processing of lexical and phrase-level prosody may proceed via different mechanisms, which may be reflected in differences in the time-course of disambiguation when the target stimulus carries compound vs. list prosody.

Given that 5-year-olds can *produce* the appropriate prosodic cues for compounds and their related list forms (Yuen et al., 2021), we predicted that children this age would also be able to use prosodic information to *identify* both forms in comprehension. Adult controls were also tested to determine if children would perform differently to adults, given that children's use of prosodic cues in production is less consistent. We further predicted that, even if these children can identify compounds vs. lists, they might show slower or different patterns in processing time compared to adults, given their inconsistent use of pauses in production (Yuen et al., 2021). On the other hand, if adult-like, we predicted that identification of both structures would be linked to the difference in boundary cues at the offset of Noun 1, i.e., the absence of boundary cues in compounds and the presence in lists. These predictions were examined using an eye-tracking method to monitor changes in looking behaviour across time as an indicator of online processing of prosodic information. We further predicted that adults, and perhaps children, might show differences in the time-course of identifying compounds vs. lists, given the differences in processing these two types of prosody found by McCauley et al. (2013). Eyetracking can thus serve as a more sensitive measure than overt measures such as pointing tasks, providing more fine-grained information on how prosodic information is processed in real time.

Method

Participants

Twenty monolingual Australian English-speaking (AusE) undergraduates (13F, 6M, 1 undisclosed) from the Sydney area formed the adult baseline for this experiment (Mean age = 21 years). An additional 6 adults were excluded for inattentiveness/low sampling rate (less than 50% samples, n = 3), or failure to calibrate (n = 3). All participants provided consent and received course credit for participation in the study.

Twenty-eight monolingual AusE-speaking children (10M, 18F) participated in the study (Mean age = 5;11 years, Range = 5;7 – 6;5). In Australia, children begin attending the first year of primary school (year 1) between 5;6 and 6;6. This age range was therefore chosen to ensure similar exposure to schooling within the sample. An additional 7 children were excluded for inattentiveness/poor sampling (less than 50% samples, n = 5), failure to calibrate (n = 1), or Autism Spectrum Disorder (n = 1). All participants received a gift card for their participation.

Stimuli

Auditory Stimuli

The stimuli were recorded in a child-friendly manner by a female native AusE-speaker in an acoustically shielded recording booth, sampled at 48 kHz. The target test stimuli were eight pairs of nouns used in both compounds and lists (see Table 1).

Table 1:

Test stimuli

Compounds	Lists
icecream	ice, cream
icecubes	ice, cubes
goldfish	gold, fish
raincoats	rain, coats
jellybeans	jelly, beans
jellyfish	jelly, fish
waterguns	water, guns
waterslides	water, slides

The targets were embedded in a carrier sentence: ‘I can see Noun1Noun2 and FILLER’ for the compound condition, and ‘I can see Noun1, Noun2 and FILLER’ for the list condition. To measure the acoustic properties of the stimuli (duration and pitch), each sentence was acoustically coded for the onset and offset of Noun1, Noun2 and Pause between the two nouns. This was to ensure that the acoustics of the stimuli were as anticipated. The onset of each noun was identified by the emergence of voicing or beginning of the release for stop consonants. The offset of each noun was identified by the cessation of voicing or frication for fricative consonants. Pause duration was measured from the offset of Noun1 to the onset of Noun2.

The average sentence duration was 2616 ms (range: 2280 to 3300ms) (see Table 2 for acoustic measures of the stimuli). The mean pitch (f0) (as measured over the voiced portion of each noun) was higher in Noun1 than Noun2 for the compounds (M_{noun1} = 264Hz and M_{noun2} = 189Hz), but similar for both nouns in the lists (M_{noun1} = 221Hz and M_{noun2} = 213Hz). No pause was found between Noun1 and Noun2 in compounds (*M* = 33ms) but a pause was present in the lists (*M* = 213ms) (see Campione and Véronis (2002) for a discussion on pauses, i.e., duration greater than 200ms, adopted as the criterion here). The average duration of Noun1 was shorter than Noun2 in compounds (M_{noun1} = 344ms and M_{noun2} = 458ms) but more similar in duration in the lists (M_{noun1} = 488ms and M_{noun2} = 497ms). These acoustic differences in pitch and duration are consistent with those reported for native speakers of Australian English (Yuen, et al., 2021).

Table 2:

Acoustic Measures of the Stimuli

	Duration (ms)		Pitch (Hz)	
	M	*SD*	*M*	*SD*
Compound				
Noun1	344	81	264	9
Pause	33	32		
Noun2	458	94	189	8
List				
Noun1	488	129	221	12
Pause	213	69		
Noun2	497	95	213	17

Visual Stimuli

Clipart pictures depicting each target sentence were created with compound and list pairs presented side-by-side, e.g., '*jellybeans and chips*' was paired with '*jelly, beans and chips*' (see Figure 1). The image associated with each sentence consisted of 3 items, e.g., red *jellybeans,* green *jellybeans* and a packet of *chips* for the compound condition, and red *jelly,* green *beans* and a packet of *chips* for the list condition. This was to ensure that the two pictures contained the same number of items and were roughly equal in complexity. A total of eight pairs of stimuli were created. Two test versions were constructed and pseudo-randomized so that no more than two trials from the same condition, e.g., two compounds or two list items, were presented sequentially. In each version participants heard auditory stimuli for each pair of images only as compound or list, never both for the same image. In each test version half of the images were matched with compounds and the other half with lists. The match was reversed to create a second test version. This ensured that each noun was heard only once, either as part of a compound or list, thereby avoiding any confusion, learning effects or predictability during the experiment. The sides (left or right) where the compound vs. list images appeared were also counterbalanced across the two versions. Half the participants received one version of the test and half received the other version.

All children included in the analysis could name all items presented. This was assessed as part of a screener that was carried out after the test session to avoid priming any prosodic cues.

Procedure

Each participant was invited into a test booth with a video feed into the control room. This was done so that the experimenter could monitor the session and parents could see their child from the control room. The participant was seated in front of a table with an LG Flatron W2753VC widescreen 27-inch high definition monitor located approximately 85cm in front of the participant. In front of the screen was a Tobii Eye-tracker X120, tilted at a 30° angle and positioned 15cm below the monitor. During the test session, participants were seated approximately 65cm in front of the eye-tracker. The experiment was delivered using Tobii Studio (3.2.3). Each trial consisted of two paired pictures, approximately 23.3cm by 27.7cm each. The auditory stimuli were played at a conversation level (≈ 65 dBA) through two external speakers on either side of the monitor.

Each session began with three orientation trials, designed to cue participants as to how to perform the task. In these trials, two pictures depicting different clipart animals were presented side-by-side on the screen (a bat and a bird, a dog and a cat, and a cow and a duck). The children were invited to play a game where they needed to look at the pictures on the computer. While viewing the images, participants heard an auditory prompt, e.g., '*I can see a bird'*, after which the target picture (e.g., bird) flashed, then both pictures disappeared.

The orientation trials were followed by four practice trials and eight test trials. Each practice and test trial lasted approximately 9.5 seconds and began with a pair of pictures. The practice trials helped children to generalise the looking response they had practiced in the orientation trials to the test condition. These pictures remained on-screen until the end of the trial. After the pictures had been displayed for 4 seconds, the auditory test sentence was played, e.g., '*I can see jellybeans and chips*'. Participants were asked to look at the picture that corresponded to the auditory sentence.

Looking Behaviour Analysis

First, areas of interest (AOIs) were defined around the compound and list pictures for each trial. Then looking behaviour in terms of fixations to the AOIs using the IVT fixation filter were extracted from Tobii Studio (3.2.3). Participants with fewer than 50% of samples recorded across trials were excluded from further analysis to avoid including data from those who were not paying attention or whose eyes could not be consistently tracked.

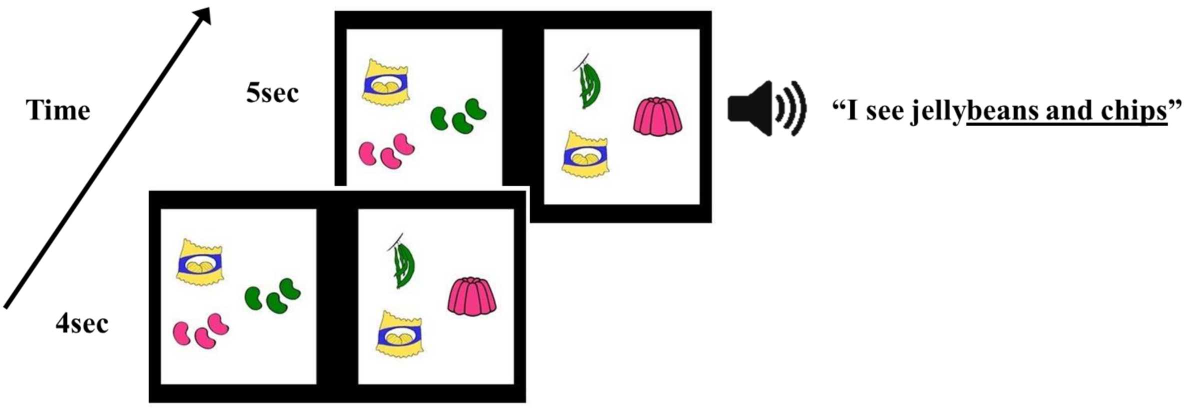

Figure 1: Sample trial with two pictures; left depicts '*jellybeans and chips*' and the competitor, right picture depicts '*jelly, beans, and chips*'. Underlined parts of the auditory stimulus are considered as the post-naming portion of the eye-tracking data (i.e., from the offset of the noun 1 in both the two-word compound and three-word list conditions).

Two types of analysis were then conducted on looking behaviour. The first examined changes in the proportion of fixations to the target picture pre- and post-auditory naming (proportion shift)[1]. The 4-second pre-naming analysis window began at the onset of each trial and ended at the onset of the auditory stimulus. The 4-second post-naming analysis window began at the offset of Noun1 in both the compound and list conditions, e.g., from the offset of *jelly* in both '*jellybeans and chips*' and '*jelly, beans and chips*'. This analysis window was selected because boundary cues are expected to begin here for list forms, but not the compounds. Fixations were binned across 200ms windows from the start of post-naming window to allow for the time taken to plan eye-movements. The 'proportion shift' measure takes participants' looking behaviour during the pre-auditory period as a within-subject

[1] Proportion shift is calculated by proportion of looks to the target picture post-naming minus pre-naming with 200ms bins.

(Proportion Shift = $\frac{Target_{post}}{Target_{post}+Distractor_{post}} - \frac{Target_{pre}}{Target_{pre}+Distractor_{pre}}$)

baseline to estimate the amount of change in looking to the target during the post-naming phase. This measure provides a gross indication of changes in looking behaviour across trials and evaluates whether children and adults show similar levels of comprehension of compounds versus lists in response to prosodic cues. Larger positive shifts (> 0) in the proportions of fixations to the target picture suggest more looks to the target post-naming and indicate correct identification of the target structure (compound or list).

The second analysis involved fitting growth curves for each participant by condition. This analysis was based on the proportion of fixations to the target over the 4-second post naming window from the offset of Noun1 in 200ms bins for both conditions. A Richards Curve (Richards, 1959) was used[2]. The parameters of the curve include estimates for the lower and upper limits of the curve; the growth rate, indicating steepness of the curve; and the time of maximum growth (inflection point), indicating time where the proportion of looks are shifting most rapidly to the target. These parameters describe a curve with an initial lag followed by a period of rapid growth before it asymptotes. Unlike the logistic curve, where the curve is always symmetrical, the Richards' Curve is more flexible in allowing for asymmetries around the inflection point. The best-fitted curve was estimated using least squared differences from the raw data. The fitted curves allow us to analyse looking changes over time, but more importantly, it allows for direct comparison of processing time (inflection point) between different conditions and groups. Shorter time to maximum growth is used here as an indicator of faster processing.

---

[2] Richards Curve ($y = A + \frac{L}{(1 \pm Te^{-k(t-tm)})^{\frac{1}{T}}}$) is a five-parameter modified logistic curve where *A* is the lower asymptote (determined by the proportion of fixations to target at offset of Noun1), *L* is the upper asymptote, *T* is a fixed inflection point, *e* is the natural exponential function, *k* is growth rate, and t_m is time at maximum growth. Note *L* is determined by the maximum of the proportion of fixations to target (a positive number) or away from the target (a negative number).

Results

To investigate whether participants had more fixations to the target after hearing the auditory stimulus, proportion shifts were compared to chance (no shift) for each Condition (compounds and lists) and Group (adults and children) separately (see Figure 2 for bar graph of results). Two sample unequal variance t-tests were conducted for each group and condition. For adults, the proportion shift was significantly above chance for both compounds (t (18) = 16.95, p < .001) and lists (t (18) = 7.72, p < .001). However, for children, the proportion shift was only significantly above chance for compounds (t (26) = 8.74, p < .001), not lists (t (26) = 0.55, p = .59). The same results remained after Bonferroni adjustments to alpha (.006). This suggests that adults identified both compounds and lists, but that children could only identify compounds based on prosodic cues.

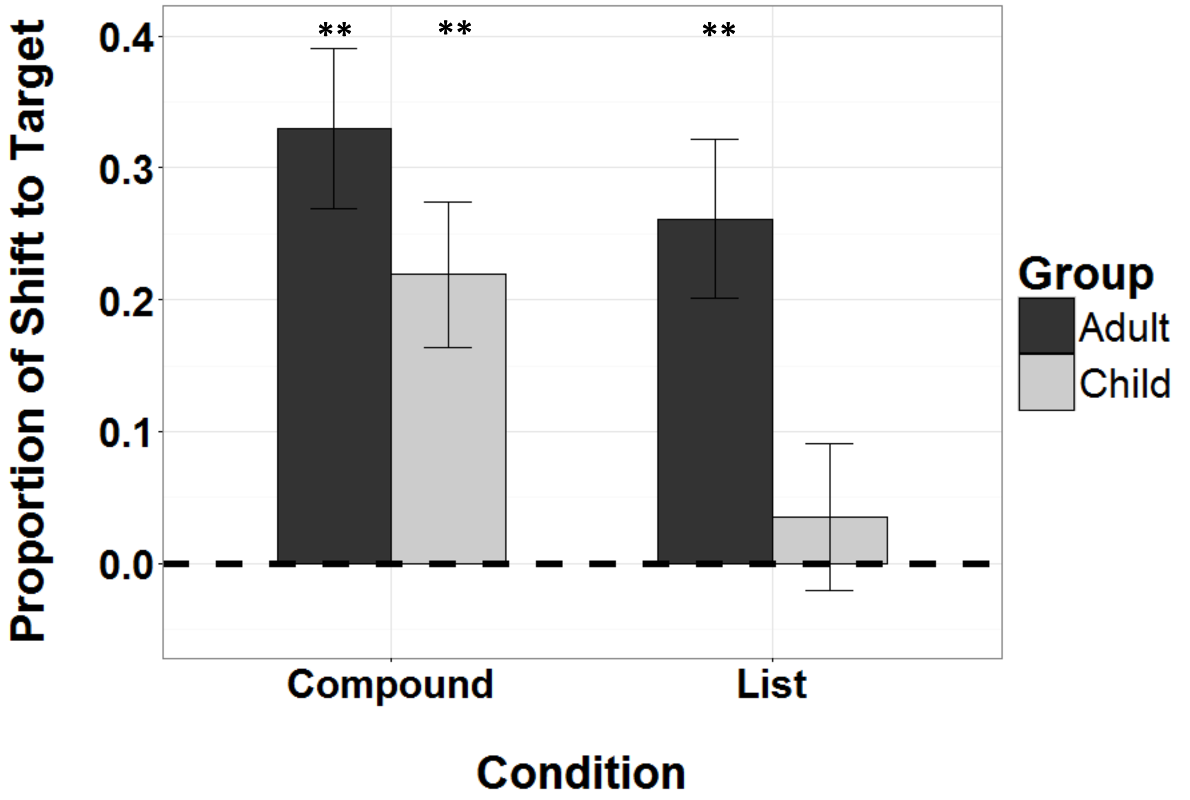

Figure 2: Proportion shift in looking to target compared to chance (0 = no shift as shown with dashed line) by Condition (Compounds and Lists) and Group (Adult and Child), with the standard error of the mean. **p < .001

To determine whether there were differences in processing time for the two conditions across groups, growth-curve analyses were performed using the curve parameter 'Time at maximum growth' as a dependent variable and indicator of processing time. This inflection point is a time when participants are beginning to make more systematic looks to the target picture across trials, and is akin to reaching a threshold point where they have accrued enough information to make a decision. Longer time to arrive at maximum growth was used to indicate longer processing time. A mixed design Analysis of Variance (ANOVA) was conducted with Condition as a within-subjects factor with 2 levels (Compounds and Lists) and Group as a between-subjects factor with 2 levels (Adults and Children) (see Figure 3, note that proportion of looks to target is around chance at the onset of Noun1 when the acoustic cues has not yet been made available to the listeners). Six outlier trials with z-scores larger than 2 were identified from the child group and were removed from the analysis. The results showed only a significant main effect for Group (F(42, 1) = 21.54, p < .001, η_p^2 = .34). This suggests that adults had significantly faster overall processing speed compared to children (M = 737ms vs. M = 1635ms, respectively) but that the two groups did not perform differently across the two conditions.

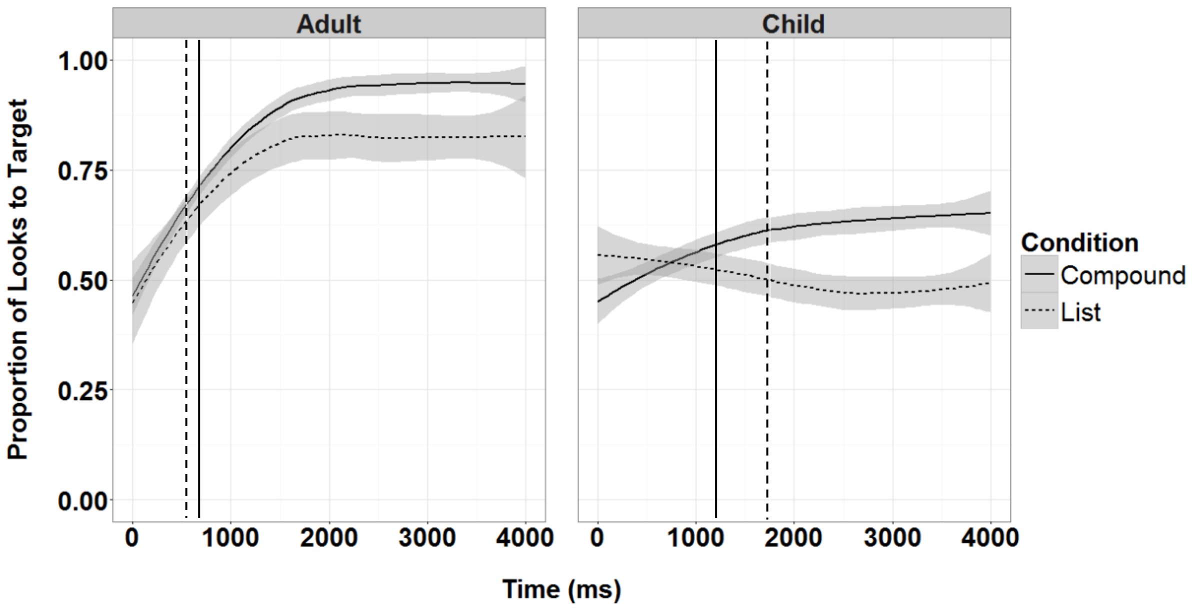

Figure 3: Raw proportion of looks to target (in 200ms bins) over time (with standard error) for compound and list conditions with 0 set to the onset of Noun1 (where relevant acoustic cues becomes available), for adults and children. Vertical lines indicate the time at maximum growth for compounds (solid line) and lists (dashed line).

Discussion

This study examined whether 5-6-year-olds can use prosodic cues to identify compounds and lists. Given that children the same age can use prosodic cues for these structures in *production*, the prediction was that children in the current study would also be able to use these prosodic cues to identify compounds and lists in *comprehension*. It was further predicted that, compared to adults, children might show slower overall processing times or different patterns in processing, especially if their productive use of boundary cues was not yet robust. Furthermore, we considered that differences in the time-course of looks across the compound and list conditions may differ within participant groups, as different mechanisms may be involved in integrating prosodic information at the word vs. phrase level of linguistic processing.

The results showed that, although adults could consistently identify both compounds and lists, the children could only reliably identify compounds; performance for lists was at chance. Therefore, prosodic cues associated with list forms do not result in more looks to the target. This suggests that they are more aware of acoustic cues associated with word-level prosody than those with phrase-level prosody. Taken together with the recent production study (Yuen et al., 2021), these results show that children can *produce* and *perceive* prosodic cues at the word level but are not yet adult-like in their use of prosodic cues at the phrase level.

In terms of the online processing of compounds vs. lists, the time at maximum growth (inflection point) of the looking data showed that adults rapidly identified compounds and lists at 795ms vs. 678ms respectively, after hearing Noun 1 in each condition. This means that adult listeners identified *compounds* and *lists* before the end of Nouns 2 (inflection points for Compounds = 795ms & Lists = 678ms (see Figure 3), end of noun 2 for Compounds = 835m & Lists = 1198ms (see Table 2)). In both cases the estimated inflection point coincides with the time when disambiguating prosodic boundary cues become available, showing that adults have fine-grained sensitivity to both the absence and presence of relevant acoustic information, using this to make their judgements. This suggests that adults can rapidly make use of acoustic information for both types of structures without incurring additional processing time.

Children, on the other hand, identify the compounds after the sentence ends, suggesting much slower processing times, or perhaps even different processing strategies compared to adults (Rigler, Farris-Trimble, Greiner, Walker, Tomblin, & McMurray, 2015). Children seemed to wait until well past the end of the sentence before making any decisions, long after the relevant time window for prosodic boundary cues had passed. This suggests that 5-6-year-olds might need much more time to process both word and phrase level prosodic cues compared to adults. On the other hand, children could not identify list forms even after all prosodic information was been made available. In compounds, word level prosodic acoustic cues are based on the strong-weak stress pattern, and shorter duration of Noun 1 than Noun 2 without pause; in lists, word stress patterns and pre-boundary lengthening of each noun, plus a pause in between, are present. However, the additional stress, durational cues and pauses did not help children identify the list forms. This suggests that children have not yet acquired the mapping between these phrase-level prosodic cues and their semantic import for signalling a list. This could also explain the inconsistent use of pauses in children's productions to contrast compounds from lists (Yuen et al., 2021). Thus, while children might be sensitive to these prosodic cues, they still need to learn to associate these cues to different prosodic (and semantic) domains. Our results thus suggest that while 5-6-year-olds can, albeit very slowly, identify word level prosodic information (compounds), they are still not able to identify phrase level prosody (lists). These findings are in line with past imaging studies which suggest that, while children may be sensitive to prosodic cues, they are not yet able to consistently evaluate and integrate this information online (Friedrich, Alter, & Kotz, 2001; McCauley et al., 2013). Future research could employ an overt decision task to see if children show the same patterns of performance found here, and/or if other factors, such as familiarity of the items or word/input frequency, might affect children's performance.

These results suggest that young children's inability to use prosodic information consistently is not due to a lexical bias, since our study and that of Yuen et al. (2021) both used real words – in contrast to Vogel and Raimy's study (2002) using non-words. Children at this age are beginning to use both word and phrase level prosody to contrast known compounds from their list forms in their own productions, albeit inconsistently (Yuen et al., 2021). The comprehension results presented here provide further support for those findings, showing that 5-6-year-olds cannot yet consistently use phrase level prosodic information in an online task to identify lists. Rather, the acquisition of phrase level prosody is more protracted than word level prosody (see Demuth & McCullough, 2009; Demuth, 2018; Gerken, 2006 for similar proposals for younger children in other domains). One reason for this might be that young children have not yet learned to weigh prosodic cues to word vs. phrase structures in an adult-like fashion, raising questions about children's attunement to prosodic information during language processing more generally.

Future studies could manipulate specific acoustic cues (e.g., pitch and duration) independently to further examine the emerging awareness of how these cues map onto different linguistics structures. It is possible that children rely on some cues more than others, leading to poor performance on the present task. Future studies could also explore the possibility that children have challenges integrating multiple prosodic cues to word vs. phrase level information, as well as the acoustic realization of compounds and lists in child- vs. adult-directed speech. This present study used child-directed speech, an input register which children are familiar with, but which might differ from adult-directed forms, e.g., slower speech rate and longer durations for segments. A better understanding of how speech style registers affect the prosodic realization of e.g. boundary cues could shed more light on cue weighting and cue use in prosodic development.

These results also raise many questions, including when children's processing of compounds and lists become more adult-like. Even less is known about how such structures are acquired by child second language learners of English, especially for native speakers of tone languages, where pitch and duration are used primarily to signal word level information (see Nguyễn & Ingram, 2007 for discussion of compound-phrasal prosody in Vietnamese). There are also implications for children with hearing loss, where pitch and duration information are transmitted differently from that experienced by children with normal hearing (Kong, Deeks, Axon, & Carlyon, 2009; Macherey & Carlyon, 2014; Macherey, Deeks, & Carlyon, 2011). A better understanding of these issues will thus be of interest for a broad range of researchers, educators and clinicians.

Conclusion

While English-speaking 5-6-year-olds can use appropriate word and phrase level prosody in production, they are only able to identify word level (i.e., compound) prosody confidently in comprehension, with the use of phrase level (i.e., list) prosody remaining a challenge. This suggests that children may be able to produce phrase level prosody before they can confidently use the same cues during language processing/comprehension.

References

Bååth, R. (2010). ChildFreq: An online tool to explore word frequencies in child language. Lucs Minor, 16, 1-6.

Becker, J. A. (1994). "Sneak-shoes", "sworders" and "nose-beards": a case study of lexical innovation. *First Language*, *14*(42–43), 195–211. https://doi.org/10.1177/014272379401404213

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *In Speech prosody 2002,* international conference.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. The Journal of the Acoustical Society of America, 95(3), 1570-1580.

Clark, E. V. (1981). Lexical innovations: How children learn to create new words. *The Child's Construction of Language*, 299–328.

Clark, E. V., Hecht, B. F., & Mulford, R. C. (1986). Coining complex compounds in English: Affixes and word order in acquisition. *Linguistics*, *24*(1), 7-29.

Clark, E. V., Gelman, S. A., & Lane, N. M. (1985). Compound Nouns and Category Structure in Young Children. *Child Development*, *56*(1), 84–94. https://doi.org/10.2307/1130176

Cutler, A., & Butterfield, S. (1990). Durational cues to word boundaries in clear speech.

Demuth, K, & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, *36*(01), 173–200. https://doi.org/10.1017/S0305000908008921

Demuth, K. (2018). Prosodic constraints on children's use of grammatical morphemes. *First Language*, 0142723717751984. https://doi.org/10.1177/0142723717751984

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. NeuroReport, 12, 3189–3191.

Gerken, L. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67–B74. https://doi.org/10.1016/j.cognition.2005.03.003

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from Japanese. *Cognition*, *136*, 350-358.

Kong, Y.-Y., Deeks, J. M., Axon, P. R., & Carlyon, R. P. (2009). Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America*, *125*(3), 1649–1657. https://doi.org/10.1121/1.3068457

Krott, A., & Nicoladis, E. (2005). Large constituent families help children parse compounds. *Journal of Child Language*, *32*(01), 139–158. https://doi.org/10.1017/S0305000904006622

Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. *Current Biology*, *24*(18), R878–R884. https://doi.org/10.1016/j.cub.2014.06.053

Macherey, O., Deeks, J. M., & Carlyon, R. P. (2011). Extending the Limits of Place and Temporal Pitch Perception in Cochlear Implant Users. *JARO: Journal of the Association for Research in Otolaryngology*, *12*(2), 233–251. https://doi.org/10.1007/s10162-010-0248-x

McCauley, S. M., Hestvik, A., & Vogel, I. (2013). Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Language and speech*, *56*(1), 23-44.

Nguyễn, A.-T. T., & Ingram, J. C. L. (2007). Acoustic and perceptual cues for compound-phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America*, *122*(3), 1746–1757. https://doi.org/10.1121/1.2747169

Nicoladis, E. (2003). What compound nouns mean to preschool children. *Brain and Language*, *84*(1), 38–49. https://doi.org/10.1016/S0093-934X(02)00519-9

Richards, F. J. (1959). A flexible growth function for empirical use. Journal of experimental Botany, 10(2), 290-301.

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental psychology*, *51*(12), 1690.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech & Hearing Research*, *37*(4), 831. https://doi.org/Article

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language*, *29*(02), 225–250. https://doi.org/10.1017/S0305000902005020

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*(4), 749-778.

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, *85*(2), B31-B41.

Wightman, C. W., Shattuck ‐ Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. https://doi.org/10.1121/1.402450

Wynne, H. S., Wheeldon, L., & Lahiri, A. (2018). Compounds, phrases and clitics in connected speech. *Journal of Memory and Language, 98*, 45-58.

Yuen, I., Xu Rattanasone, N., Schmidt, E., McDonald, G., Holt, R., & Demuth, K. (2020). Five-year-olds produce prosodic cues to distinguish compounds from lists in Australian English. *Journal of Child Language*, 1-19.