Abstract

Uptalk refers to the use of rising intonation on declarative utterances. Previous research has shown that, at age 6 years, children use rising contours with declaratives more frequently than adults, and this pattern appears to persist until 14 years of age. However, it is unclear why such a trend persists. To gain a clearer developmental picture of uptalk, the present study analyzed the form and function of uptalk produced by children aged 6 to 7 and to 11 years from the American Midwest, using a storytelling task. Contrary to previous findings, the results indicate that children of both age groups use uptalk in an adult-like way: they overwhelmingly favor L-H% over H-H% boundary tones, and most strongly associate the contour with continuation. The lack of age differences suggests that children’s use of uptalk is comparable to that of adults by the age of 6, at least in certain narrative contexts. The use of a familiar storytelling task in the current study may explain the greater success observed for children than in previous studies, suggesting the relative importance of the elicitation task in the investigation of child speech.

Index Terms: uptalk, rising intonation, child speech, intonation development, prosodic acquisition

1. Introduction

Some aspects of intonation are not fully mastered until the pre-teen years [1, 2]. Such protracted development can be explained by intonation being a complex system that interacts with various linguistic domains, including grammar, pragmatics, and social attitudes. Intonation is assumed to develop with cognitive maturity and linguistic knowledge, with increasing age [1, 2].

Uptalk, the phenomenon in which a speaker produces rising intonation at the end of a declarative utterance, is a common feature of adult speech across varieties of English [3]. Despite the prevalence of uptalk, however, little is known about its development due to the relative scarcity of research on children’s use of uptalk. The goal of the present study was to examine the use of uptalk in narratives produced by child and adult speakers from the American Midwest.

The prosodic form of uptalk involves an f0 contour with a final rise. Under the phonological framework of Tones and Break Indices of Mainstream American English (ToBI-MAE) [4], four uptalk contours have been attested in varieties of North American English: H*-H-H%, L*-H-H%, H*-L-H%, and L*-L-H% [5, 6, 7]. However, different varieties show preferences for different contours [5, 6, 7]. For example, in Midwestern American English, speakers rarely use the H*-H-H% contour for uptalk in read speech [5], while the L*-L-H% contour is the most preferred form for uptalk in Southern Californian English, followed by the H* H-H% contour [6]. Conversely, the L* H-H% and H* L-H% contours are used for uptalk in South Ontario English [7].

The use of uptalk is dependent on the discourse and interactional context in which it occurs [6, 8, 9]. For example, uptalk is found most often in narratives compared to other discourse contexts. This higher likelihood of uptalk in a narrative is likely due to the high semantic complexity of narratives, which requires speakers to form structural coherence; uptalk can be used as a cohesive device to signal continuation [7, 10], for example. Further, although uptalk is found in interactional contexts, for example, as a floor-holding device, it is also observed in less interactional contexts, such as read speech [5]. In read speech, uptalk can be used to convey uncertainty about what is being said [11], but it can also be used to make assertions or convey information, a function that is typically associated with a falling contour in English [6, 12].

Children’s use of uptalk has been investigated in several previous studies with different age groups. Before age 5, children exhibit difficulties in producing and imitating rising contours reliably [13, 14]. By age 6, however, children appear to reliably produce rising contours [15]. In fact, starting at age 6, they start to use rising contours on declaratives more than adults [16]. This tendency appears to persist for many years and is even found in children as old as 14 years [17]. Moreover, for younger children, the choice of tunes may differ from that of adults. Armstrong et al. [16] acoustically analyzed the form of rises and found that 6-year-olds use shallower rises than adults, employing steeper rises for questions and shallower ones for non-questions. Wolff [17], on the other hand, observed that the form of rises produced by an 11-year-old girl and a male adult was very similar. We therefore considered both the frequency and form of rising contours in the current study.

One hypothesis about the apparent overuse of rising contours is that children might not have yet acquired the various functions associated with rises and their context-appropriate use. Armstrong et al. [16], for example, observed that children use rising contours more mid-narrative than adults, while reserving falling contours more for the conclusion. They proposed that this tendency might indicate children’s more limited knowledge of the functions of intonational contours compared to adults’. Similarly, Wolff’s [17] analysis suggested that functions of uptalk may be different across ages. Although there was no consensus among the three raters on the interpretations of the functions, it was consistent across the raters that the child used uptalk for the typically reported functions, such as the need for reassurance, whereas the adult used uptalk for functions that have not been reported previously, including emphasis.

We propose an alternative hypothesis to explain children’s different use of rising contours from that of adults in previous work [16, 17]. Specifically, children’s ability to use rising contours like adults may be limited by the speech context. That is, the ease and familiarity of a speech context may affect children’s performance in demonstrating their uptalk skill. To test this hypothesis and to better understand the development of uptalk, the present study analyzed the form and function of
rising contours produced by 6- to 7-year-old girls, 10- to 11-year-old girls, and female adults from the American Midwest, using a storytelling task guided by wordless picture books of familiar fairytales.

In previous studies [16, 17], personal narratives were elicited by a female adult experimenter by asking questions about the participants. While such elicitation tasks may not appear too difficult, the task in the current study may be easier in two respects. First, participants may expend less cognitive effort in planning narratives and thus focus more on the task of storytelling, since they can follow a plot that is predetermined and provided by visual cues [18]. There is no such contextualization when telling personal narratives, requiring speakers to rely on memory, making the task cognitively more demanding, particularly for younger children [18, 19]. Second, less social interaction with an unfamiliar female experimenter in the current study may have enabled children to perform better. As noted by Ainsworth [20] and Wolff [17], talking to an unfamiliar adult may influence the extent to which children demonstrate their uptalk skill. For example, the adult participants might have found the adult experimenter more comfortable and more their equal, while the children might have found talking to an unfamiliar adult uncomfortable and felt less secure. Additionally, the storytelling task used in this study may provide a more useful context for examining child speech as a limited set of functions are expected to be conveyed by uptalk in this context, allowing for a more focused analysis.

If children’s overuse of rising contours is due to a protracted developmental trajectory, given the complexity of intonational meaning in English, we expected to find children’s more frequent use of rising contours than adults and associations between form and function that are different from those of adults. However, if children as young as 6 years old have acquired adult-like use of uptalk, and if it was the difficulty of the speech context that prevented them from showing their competence in previous work, we expected that children would use rising contours like adults in both form and function, in the more familiar storytelling task in the current study.

2. Methods

2.1. Materials

The data used in this study were collected in a laboratory in a science museum in the American Midwest as part of a larger project involving speech production and perception by children and adults [21]. Production data of a subset of participants who met the following criteria were analyzed: female monolingual, native speakers of Midwestern American English belonging to the age groups of interest. This selection process yielded a total of 24 participants from three age groups: 8 girls aged 6-7 years (M=6.8), 8 girls aged 10-11 years (M=10.6), and 8 female adults (19-29 years, M=22.5).

The data were produced in a storytelling task, in which participants were asked to narrate wordless picture books of two well-known children’s stories, “Little Red Riding Hood” [22] and “Goldilocks and the Three Bears” [23], in their own words. “Little Red Riding Hood” consisted of ten pictures and “Goldilocks and the Three Bears” consisted of eleven pictures. Given that uptalk use is highly context-sensitive [8, 9, 10], having the same set of pictures and stories make the data comparable across speakers, while eliciting natural speech.

Each picture was shown on a computer screen one at a time, controlled by E-Prime. Participants saw the entire set of pictures for each story twice, once to familiarize themselves with the story and then a second time to narrate. Participants narrated the story at their own pace and clicked the mouse to proceed to the next picture. All of the narratives were recorded using a head-mounted microphone (Audio-Technica ATH-770COM) and the Audacity recording software at a 44.1 kHz sampling rate. The order of the stories was randomly assigned.

2.2. Prosodic annotation

The narrations of both stories produced by each participant were analyzed using ToBI-MAE [4]. The precise status of nuclear pitch accents occurring at the right edge of an intonational phrase (IP) boundary is unclear in this variety, as they may not be functionally distinctive for Midwestern American English speakers [24]. Thus, we leave the effects of pitch accents on uptalk in this variety for future work. Instead, this study focuses on edge tones, or combinations of a phrase accent (L-H-) and a boundary tone (H%): L-H% and H-H%. Contours at the right edge of IP boundaries, which include the phrase-final pitch accent, the phrase accent, and the boundary tone, were coded as either rising or non-rising, based on their annotated edge tones. IPs with L-H% and H-H% edge tones were coded as rising, regardless of pitch accent; bital phal pitch accents (e.g., L-H+1) as well as monotonal pitch accents followed by these edge tones were coded as rising. IPs with other edge tones were coded as non-rising.

All data were annotated by the first author. To assess intra-coder reliability, a subset (25%; 2 of 8 speakers from each age group) of the data was randomly chosen and was reannotated by the same coder six months after the first coding. Cohen’s kappa coefficient (κ) was computed using the irr package [25] in R (Version 3.6.1) [26]. The intra-coder reliability was high, with the consistency of 96% (κ = 0.94, p < 0.001). The first set of annotations was used for analysis.

IPs were excluded from the analysis if they were not part of the narrative, such as repeating the instructions. A total of 1,424 IP tokens were included in the analysis.

2.3. Function annotation

The IPs were coded for their function by two raters based on both text transcripts and listening. Three uptalk functions that are relevant to the task in the current study were considered: continuation, uncertainty, and statement. Continuation was defined as linking phrases together or signalling incompleteness [7, 10]. An IP token was coded as continuation if a subsequent utterance showed the same subject omission and/or if it was followed by a relatively short pause (<500ms) or a conjunction (e.g., and, but). Uncertainty was defined as showing hesitance or lack of confidence in either the accuracy or acceptability of what was being said [11]. This function was often accompanied by signs of disfluency (e.g., pauses, lengthening). Statement was defined as making an assertion or conveying information [6, 12]. An IP token was coded as statement if it was followed by a relatively long pause (>500ms) and/or if it was the last item of a list.

List and question status were also included to distinguish these functions from uptalk. List was defined as listing of more than two items in a row. Question was defined based on the syntactic structure of the IP, such as subject-auxiliary inversion or wh-questions. Of the eight questions that were produced, six were wh-questions and had a falling contour. The other two questions were marked by the H-H% edge tone (see Figure 2). There were no IP tokens that these five function categories
All data were coded by the first author and a trained research assistant, following the criteria outlined above. The second coder is a native speaker of Midwestern American English and was blind to the phonological annotation. 95% (1,353 out of 1,424 tokens) of the coded data was consistent between the two coders. Cohen’s kappa coefficient (κ) between the two coders was 0.92 (p<.001), indicating high inter-coder reliability. Given the high agreement score, the data coded by the first author were used for analysis.

3. Results

On average, each speaker produced 33 IP tokens in the younger child group, 26 IP tokens in the older child group, and 29 IP tokens in the adult group across the two stories. The distribution of the functions of these IP tokens is shown in Table 1. Children, both younger and older, produced continuation most frequently, while adults produced statement and question equally frequently. For speakers in all age groups, uncertainty and question functions were rarely produced. Thirteen distinct contours were identified from the phonological analysis. Rising contours with L-H% and H-H% edge tones accounted for approximately 28% of all IP tokens produced across age groups.

Table 1: Number of IP tokens for each function in each age group.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6-7 years</td>
<td>239</td>
<td>13</td>
<td>134</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td>10-11 years</td>
<td>242</td>
<td>1</td>
<td>173</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>Adults</td>
<td>198</td>
<td>3</td>
<td>198</td>
<td>75</td>
<td>6</td>
</tr>
</tbody>
</table>

3.1. Use of rising contours

Figure 1 shows the mean proportion of rising contours relative to all other contours for each function in each age group.

To explore the effects of Age Group (6-7 year-olds, 10-11 year-olds, adults) and Function (uptalk functions: continuation, statement; non-uptalk function: list) on the use of rising contours (rising or non-rising), a logistic mixed-effects model with the maximal random-effects structure that converged was constructed in R using lme4 [27]. Uncertainty and question functions were not included in the analysis as there were too few tokens of these function types (see Table 1). Random effects included random intercepts by picture and participant. Log-likelihood model comparisons were used to assess statistical significance. Post-hoc pairwise comparisons of significant effects were conducted using Tukey’s HSD tests in the emmeans package [28].

The proportion of rising contours was affected by Function ($\chi^2(2)=43.16$, $p<.001$), but it was not affected by Age Group or its interaction with Function. The pairwise comparisons for Function revealed that speakers across age groups using rising contours significantly more often for continuation compared to statement ($\beta=0.97$, $SE=0.16$, $z=6.16$, $p<.0001$) and list ($\beta=0.62$, $SE=0.23$, $z=2.72$, $p=.02$). The pairwise comparison for statement and list was not significant. These results suggest that a rising contour is most strongly associated with continuation in Midwestern American English. The lack of an Age Group effect or interaction suggests that children in both age groups used rising contours at the same rate and for the same functions as adults from the same dialect region when narrating a story.

3.2. L-H% vs. H-H%

Figure 2 shows a summary of the mean proportion of L-H% edge tones within each function and age group, out of all of the rising contours (i.e., those with L-H% and H-H% edge tones). Across all age groups, except for uncertainty and question, L-H% was overwhelmingly preferred to H-H%.

![Figure 2: Mean proportion of L-H% for each function in each age group, with standard errors.](image)

To explore the effects of Age Group and Function on the use of the two rising edge tones (L-H% or H-H%), a logistic mixed-effects model with the same fixed effects (Function and Age Group) was built. The converged model included a random intercept by participants. The main effect of Function on the use of the edge tones was statistically significant ($\chi^2(2)=10.16$, $p=.006$). The effect of Age Group and its interaction with Function were not significant. In the pairwise comparisons, the L-H% contour was produced marginally more often with continuation than with list ($\beta=1.29$, $SE=0.56$, $z=2.32$, $p=.053$). None of the other pairwise comparisons were significant. Although rises are often used with the list function in these data, given that H* H-L% is traditionally considered “list intonation” in English [29], the modestly less frequent use of the L-H% contour with lists in these data is not surprising. The lack of an Age Group effect or interaction suggests that the frequency and distribution of the two edge tones did not differ across age groups.

4. Discussion

In the current study, a storytelling task was used to investigate the use of uptalk by children aged 6 to 7 years and 10 to 11 years, and by adults, speaking Midwestern American English. There were no age differences in the use of uptalk among the children of both age groups and adults. Across age groups, rising contours on average accounted for approximately 28% of the total utterances produced. Further, speakers in all age groups associated rising contours most strongly with...
continuation and showed a preference for the L-H% contour (90%) relative to the H-H% contour (10%). The use of uptalk by the children in both age groups in the current study was therefore comparable to that of the adults in terms of frequency, form, and function. The overall lack of age differences in the present study is inconsistent with previous findings [16, 17], in which the overuse of rising contours on declarative utterances was found among children aged 6 to 14 years. However, a larger sample might reveal a significant interaction between age and function (see Figure 1).

Two hypotheses were proposed to explain why children as young as 6 years old overuse rising contours on declarative utterances compared to adults speaking the same variety. One hypothesis, suggested in previous research [16, 17], attributes the children’s behavior to a protracted developmental trajectory due to the complexity of intonational meaning in English, whereas the other hypothesis ascribes it not to children’s immaturity in using uptalk but to differences in the difficulty of the task used to elicit narratives from children.

Together, the results of the current study appear to support the hypothesis that the difficulty of the task that is used to investigate children’s ability to use uptalk can influence the extent to which children can demonstrate their knowledge and skills in this domain. Therefore, it is possible that children in previous research failed to use uptalk like adults not because they had not yet acquired its appropriate usage, but because the speech context made it harder for them to demonstrate their competence.

In this study, participants were asked to narrate well-known children’s stories in reference to a sequence of pictures that were presented on a computer screen. In this context, a limited set of functions, as examined in this study, are expected. Further, the cognitive load may be reduced compared to the personal narrative context used in previous research because the content and structural support is provided with visual cues [18, 19]. The relative ease of the task may have enabled the children to better demonstrate their skill with uptalk. In Armstrong et al.’s study [16], where girls aged 6 to 7 years were found to produce non-question rises more than their mothers, participants were asked to tell a story about their own personal experiences. While the content of a personal narrative is something that speakers have knowledge about, the amount of effort required to plan and tell a story about themselves to a stranger could be greater than the effort required to tell a familiar story guided by a sequence of pictures. This greater demand may leave fewer processing resources available for specific tonal choices.

Moreover, the interactional environment could have affected the children’s choice of tunes. The less interactional environment could have increased the likelihood that children would successfully mirror adult patterns of uptalk. That is, the relatively limited interaction with an experimenter may have allowed children to focus more on the task, as they could allocate less attention to the monitoring and management of social interactions, leading to less overall cognitive demand. In both Armstrong et al.’s [16] and Wolff’s [17] studies, personal narratives were elicited from children and adults in an interview with a female adult experimenter. It is possible that there were differences in the status of speakers in the interview context between adult and child participants, which could affect children’s use of uptalk, as suggested by McLemore [9], Wolff [17], and Ainsworth [20]. As noted above, the adult participants might have found interacting with the adult experimenter more comfortable than the children. Such differences could have either hindered children’s performance or led the children to use non-question rises more frequently to achieve other functions that would not be used by adults in that context. For instance, they might have felt like they needed to seek their listener’s verification [20] or show deference [11].

Thus, the more familiar storytelling task in the current study may have enabled us to observe the mastery of uptalk form and certain functions at an earlier age compared to previous research. This earlier success suggests the importance of the task when investigating child speech, as children may be more vulnerable to task difficulty and other aspects of experimental design than adults [20]. For instance, Ainsworth [20] partially attributed the overall rarity of uptalk in the 4-year-olds’ speech in her study to her unsuccessful elicitation of speech in a more relaxed environment and to their potentially limited narrative skills. However, it should also be noted that the limited interaction involved in the storytelling task constrained our ability to examine other, more interactive functions associated with uptalk, such as checking listener’s comprehension [8], which might allow us to observe the development of uptalk in relation to children’s social competence. Further, while the relatively easier elicitation task may benefit us in understanding younger children’s use of uptalk, the relative difficulty of a task might not account for the overuse of uptalk by children as old as 14 years in previous work [17]. However, potential differences in the status of the speakers in interactional contexts might still provide some explanation for the overuse of uptalk by teenagers. A functional analysis of uptalk use in interactional contexts across age groups is therefore necessary to gain a better understanding of the development of other uptalk functions that were not examined in the current storytelling task.

In addition to the comparable distribution of rising contours by function across age groups, children aged 6 to 7 and 10 to 11 years also showed the same preference for uptalk forms as adults in the current study. Although our results are not directly comparable to previous studies due to different methods of formal analysis of rising intonation, all child and adult speakers in this study preferred the L-H% contours for uptalk to the H-H% contours. Thus, unlike children who used shallower rises than adults in Armstrong et al.’s study [16], all of the participants in our study preferred a shallower phonological rise for uptalk. This result may reflect dialect variation in tonal inventory across varieties of North American English. Alternatively, the prosodic annotations used in the present study might not have captured more subtle differences that were observed in the previous acoustic analyses. Future research is necessary to confirm such possibilities by employing both phonological and acoustic analyses to examine the form of rises across North American English varieties.

Taken together, the analysis of phonological forms of IP contours, combined with the function analysis in this study allowed us to observe that children as young as 6 years old use uptalk with a similar overall frequency, for similar functions, and with the same phonological form as adults from the same local dialect region in a picture-guided storytelling task.

5. Acknowledgements

We would like to thank Elizabeth Bohinski for helping with the function annotation.
6. References


