10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Phonological and Phonetic Realizations of Downstepping in Child Speech

Jill Thorson[1], *Stefanie Shattuck-Hufnagel*[2]

[1]University of New Hampshire, USA
[2]RLE, MIT, USA

jill.thorson@unh.edu, sshuf@mit.edu

Abstract

Recent research has applied an autosegmental-metrical approach to child speech in Dutch, Portuguese, Spanish, and Catalan, and there is an ongoing need to determine how this model and the ToBI transcription system can be used to represent the speech of young children. Specifically, the role of downstepping in child speech has not yet been analyzed in detail. The motivation for this study is to examine the phonetic implementation of the downstepped H+!H* pitch accent, as well as begin to create an inventory of the intonational patterns found in American English-speaking toddlers.

Phonological and phonetic analyses were carried out for utterances from two corpora of spontaneous child speech, elicited during a storybook task and a game-based interaction. Phonological results show that H* and H+!H* are the most frequently occurring pitch accents, with other types appearing less frequently (L*, L+H*, L*+H). Phonetic analyses of H+!H* tokens show that its phonetic implementation is varied, with at least six sub types identified when produced with a low boundary tone. Taken together, the results begin to form a picture of the sophisticated intonational patterns present in the speech of young children. Further research is necessary to determine the connection between the phonological and phonetic representations.

Index Terms: child language acquisition, intonational phonology, autosegmental-metrical theory

1. Introduction

The interaction between the phonological category of a pitch accent and its phonetic implementation is not always straightforward. In the case of downstepping, this connection has been more closely linked to the phonetic level than the phonological one ([1]). In the autosegmental-metrical approach and the ToBI transcription system, downstepping provides the speaker a mechanism to target tones that are not restricted to only two-levels (low, L and high, H), but also tonal targets that fall between these two points ([2]–[4]). For example, the !H* pitch accent is most often used to represent the gradually lowering fundamental frequency (F0) contour on consecutively accented words in a single utterance. A second downstepped pitch accent is the H+!H*, a notational variation to the H+L* pitch accent emerged in the original ToBI system. To date, this pitch accent has rarely been used to describe child speech. The motivation for this study is to analyze both the presence of and the phonetic implementation of the downstepped H+!H* pitch accent in the speech of toddlers in American English.

The H+!H* is unique as it is the only bitonal pitch accent without contrasting low-high tonal targets. Instead, this pitch accent has two high targets, with the second one downstepped from the first. Unlike the other downstepped varieties, this pitch accent does not require the presence of a preceding accented word in order to necessitate the downstep. Rather, the H+!H* is notably characterized by a high preceding tone onto a downstepped high prominent tone on the accented syllable ([5]). This allows the H+!H* to occur in an utterance with only a single pitch accent. This is critical when considering child speech where many early utterances contain only a single prominent or accented element.

In the adult literature, limited research has been conducted regarding the use and implementation of the H+!H* pitch accent. In the original ToBI transcription system, this pitch accent is represented as H+L*, with little evidence for an associated meaning. Originally, it was postulated to indicate an "instantiation with prediction" ([6], p. 298). The current ToBI training materials for Mainstream American English present the phonetic implementation of the H+!H* pitch accent and note its occurrence in an utterance such as "*You want another example? Mother Theresa*", with the H+!H* occurring on '*Theresa'* as a falling F0 contour across the word ([5]). The meaning is not overtly identified, but is minimally one of referential newness. In an analysis of American English spoken radio corpora, the H+!H* pitch accent is rare, appearing in only 5% of the data (versus the medial F0 peak H*, which accounted for 90% of the data) ([7]).

Relatedly, in German, the H+!H* has also appeared in research looking at read speech, where the presence of this pitch accent has been found on new and accessible referents ([8]). Additionally, there has also been debates in German on whether the H+!H* may be phonetically and semantically distinct from the H+L* pitch accent ([9], [10]). Taken together with work in English, the H+!H* downstepped pitch accent is phonetically classified by its preceding high tone and its fall over the course of the stressed syllable. Questions still remain about whether this is its only phonetic implementation, particularly when produced in a more spontaneous speech setting. Additionally, research has been limited to adult speech, while little is known about this pitch accent in the speech of young children.

Several studies have successfully applied an AM approach to child speech (Chen & Fikkert, 2007 [11] for Dutch; [12] for Portuguese; [13], [14] for Catalan and Spanish). For example, research in Catalan and Spanish shows that by two-years-old, child speech consists of the basic intonational categories ([14]), including contours that would not be predicted by biological constraints. Previous work in American English has also shown that toddlers are able to manipulate the acoustics in a sophisticated manner to produce a range of pitch accents ([15]). Still, little research has looked at the range of pitch accents present in the speech of young children in English or the phonetic variations that may exist amongst these pitch accents. Most work has focused on a phonological analysis, with less detail on how the pitch accents are represented phonetically in child speech.

10.21437/SpeechProsody.2020-69

The primary aim of this study is to analyze the phonological and phonetic realizations of the downstepped H+!H* pitch accent in child speech. This aim is two-fold in that we will examine (1) the distribution pattern of H+!H*, and (2) the variety of phonetic manifestations of H+!H* in the spontaneous speech of toddlers. A secondary aim is to begin to create an inventory of pitch accents and boundary tones produced by American English-acquiring toddlers, using the ToBI transcription system and an AM approach.

2. Method

Data were analyzed from two child speech data sets that consisted of structured spontaneous speech interactions. A phonological analysis was conducted and a subset of the data was additionally analyzed phonetically.

2.1. Participants

Child speech data were analyzed from the Imbrie speech data set from MIT ([16]) and the Thorson speech data set from Brown University ([15]). Data from eight participants were analyzed (4 female) with a mean age of 2;7 (*range* = 2;5 to 2;2). All participants were from the New England region of the USA. The Imbrie participants were from the Boston, Massachusetts, vicinity and the Thorson participants from the Providence, Rhode Island, vicinity. All participants were typically developing with no known visual or hearing impairment, and no reported speech, language, or developmental disorders. Participants received a small monetary compensation or gift for their participation.

2.2. Procedure

Both the Imbrie and Thorson studies collected speech data from children in naturalistic settings. The Imbrie task asked children to describe images in a story book to an experimenter. There were target referents present throughout the story in order to elicit multiple repetitions of the same word (e.g., *'mud'*). The Thorson task engaged children in a game with the experimenter where the same items were present for each participant (e.g., *'moon'*). Each study elicited the production of a set of target nouns that were consistent throughout each of the studies. Critically, both studies collected spontaneous speech data from the children, with the tasks providing a structure to the interactions but allowing children to produce utterances in a natural manner. Both data sets were designed to collect speech from young children, targeting the early period of language development.

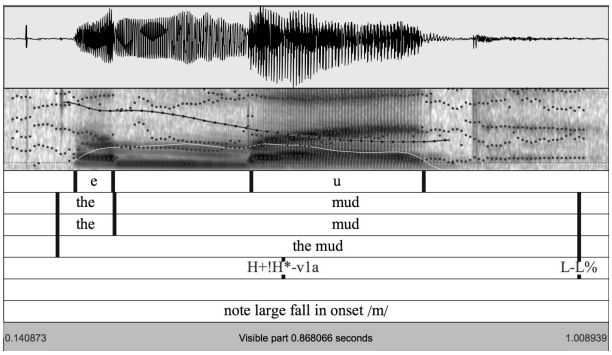


Figure 1: *Example waveform, spectrogram, pitch contour, and textgrid for the utterance 'the mud'. Visible pitch range for pitch contour: 50-500 Hz.*

For the current analysis, child utterances were first extracted as wave files from the full interactions to be analyzed individually (Figure 1). Partially and fully unintelligible utterances were excluded from the analysis.

2.3. Analyses

Phonological and phonetic analyses were conducted to examine the H+!H* pitch accent and any of its variations, and also to determine overall pitch accent distributions. Using Praat, a textgrid was created for each wave file ([17]). Each textgrid consisted of seven tiers: (1) vowels, (2) syllables, (3) words, (4) utterances, (5) tones, (6) alternate tones, and (7) comments. Tiers (1)-(4) were interval tiers marking the onsets and offsets of the designated tier (e.g., words). Tiers (5) and (6) were point tiers and were used to annotate the pitch accents and boundary tones for each utterance. Tier (7) allowed for comments by the annotators. See Figure 1 for an example waveform, spectrogram, and textgrid for the utterance *'the mud'*.

2.3.1. Phonological Analysis

In Praat, utterances were labeled following the ToBI annotation system for Mainstream American English (MAE_ToBI; [5]). Table 1 shows a breakdown for the total number of utterances and pitch accents annotated by participant and overall. A total of 310 utterances were analyzed across both data sets with 439 pitch accents annotated.

Table 1: *Participant ID, age, sex, number of utterances transcribed, and total number of pitch accents annotated. In the ID, T denotes the Thorson data set; I denotes the Imbrie data set.*

ID	Age	Sex	# of Utterances Analyzed	# of Pitch Accents Annotated
T01	2;5	M	21	27
T02	2;6	M	47	63
T03	2;6	F	57	71
T08	2;7	M	112	158
I01	2;6	F	37	70
I02	2;7	F	8	14
I03	2;8	M	7	11
I08	3;2	F	18	25
TOTAL			**310**	**439**

2.3.2. Phonetic Analysis

Two phonetic analyses were performed. First, for the H+!H* pitch accent, additional notation was made based on the various phonetic realizations that were encountered across utterances. The different variations of H+!H* are noted in the phonetic results section. Second, a subset of the data was selected in order to conduct a more in-depth acoustic analysis between the H* and H+!H* pitch accents. The subset included only the word 'mud' from the Imbrie data set, which included many occurrences of this word due to its methodological design. There were 57 instances of the word 'mud' between participants I01 (30 instances) and I02 (27 instances).

For each occurrence of 'mud', a set of acoustic landmarks for the H* and H+!H* pitch accents were marked. For the H* pitch accent, the F0 was marked at the onset of the accented syllable, the peak of the pitch accent, and the offset of the rise in the accented syllable. For the H+!H* pitch accent, the F0

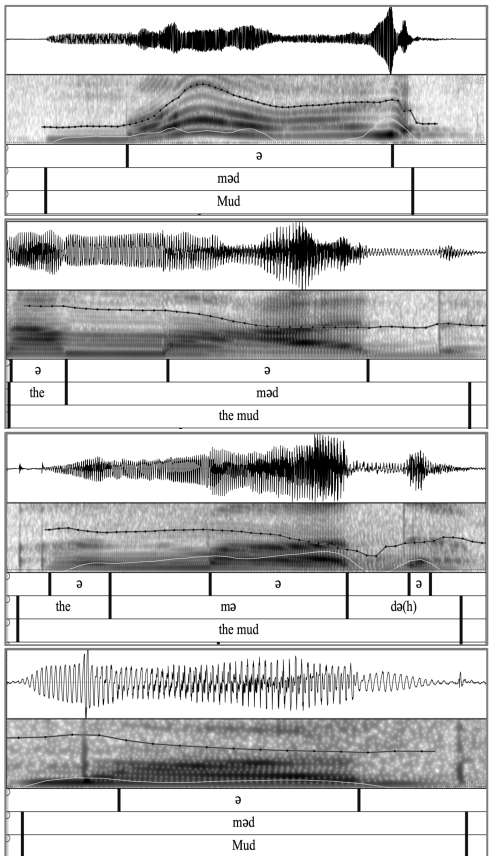


Figure 4: *Example 'mud' tokens for H*, H+!H*v1a, H+!H*v2b, and H+!H*a-np (from top to bottom).*

peak of the preceding H tone was marked as well as the F0 at the accented syllable onset and offset. Two F0 difference scores were calculated for each pitch accent type using the F0 values from these points (σ = syllable).

- *Difference 1-H**: F0 high peak – F0 σ onset
- *Difference 2-H**: F0 σ low – F0 high peak
- *Difference 1-H+!H**: F0 σ onset – F0 previous high
- *Difference 2-H+!H**: F0 σ offset/L – F0 σ onset

3. Results

Phonological and phonetic results are presented for the data from the two child speech data sets.

3.1. Phonological Results

For the overall pitch accent distributions, all children produced the H* pitch accent. All but one of the participants showed some type of instance of the H+!H* pitch accent (T01 did not have any occurrences). The L* pitch accent was the next most frequent pitch accent after the H* and H+!H*. Other pitch accents that were annotated include L*+H, L+H*, and !H*. For boundary tones, L-L%, L-H%, H-L%, and !H-L% were annotated. The L-L% boundary tone was the most frequent for these data sets. Figure 2 shows the percentage of occurrences of each pitch accent by participant and overall. Overall, the H* pitch accent was the most frequent (55%), then the H+!H* (21%), the L* (12%), and finally the 'other' category (12%).

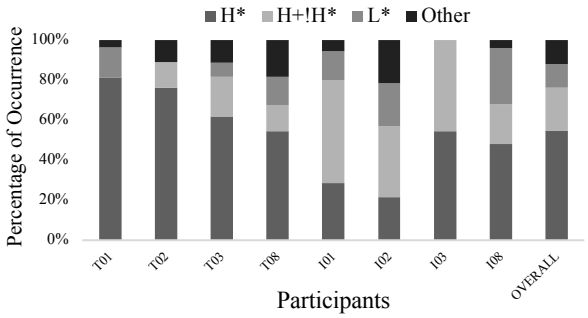


Figure 2: *Pitch accent distributions for each participant and overall. The 'other' pitch accent category includes !H*, L*+H, and L+H*.*

A further analysis of the sub-types of the H+!H* pitch accent was conducted to reveal three primary versions and two types. The details for each type are described in the phonetic analysis with Figure 3 providing the tonal schematics and descriptions for these versions when produced with a L-L% boundary tone. We used the H* pitch accent as a baseline to compare the H+!H* variations, which consists of a rise on the accented syllable. While there was no rise on the accented syllable for H+!H*, several varieties were encountered.

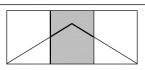
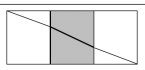
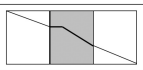
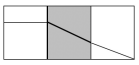
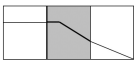
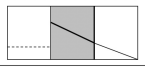
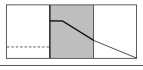
Pitch Accent & Boundary Tone	Tonal Schematic *Shaded region = accented syllable*		Description *σ = syllable*
	Type a *steep f0 fall at onset of accented σ*	Type b *f0 plateau at beginning of accented σ followed by a fall*	
H* L-L%		N/A	
H+!H* version 1 L-L%			Higher f0 on preceding σ
H+!H* version 2 L-L%			f0 is the same height on the preceding σ as it is on accented σ
H+!H* version 3 L-L%			No preceding σ for f0 (**np**) OR low f0 on preceding σ (**p**)

Figure 3: *Tonal schematics and descriptions for H* and H+!H* pitch accents low boundary tones.*

3.2. Phonetic Results

The first phonetic analysis examined the different phonetic realizations of the H+!H* pitch accent in child speech. Figure 3 shows the three versions and two types identified when occurring with a low boundary tone. The three *versions* describe F0 before the accented syllable, while the two *types* describe F0 during the accented syllable. Version 1 consists of a higher F0 on the preceding syllable, version 2 has an F0 at the same level on the preceding syllable as it is at the onset of the accented syllable, and version 3 has either no preceding syllable for F0 to appear (np) or it occurs low on the preceding syllable (p). Type a consists of a steep fall at the onset of the accented syllable and type b consists of a partial F0 plateau at the beginning of the accented syllable followed by a fall. Figure 4 provides four examples of the word 'mud' for the versions and types encountered.

The second phonetic analysis analyzed the acoustic parameters of H* and versions 1 and 2 of the H+!H* pitch accents (types a and b collapsed). Figure 4 shows the change in F0 leading up to the accented syllable as well as the change in

F0 during the accented syllable for the H+!H* varieties. Difference 1 shows that while H* demonstrates a rise, the H+!H* version 1 shows a fall, and the H+!H* version 2 shows the F0 plateau (Figure 5).

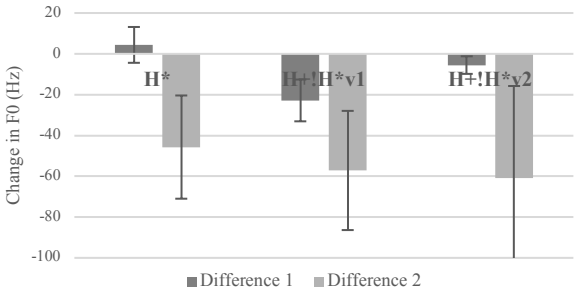


Figure 5: *Change in F0 for H*, H+!H*v1, and H+!H*2 (v3 is not included as it does not typically occur with a preceding syllable).*

4. Discussion

This primary goal of this study was to examine downstepping in the spontaneous speech of young children. Using an AM approach and the ToBI annotation system, the pitch accent distributions for eight toddlers were analyzed. The phonological results show that the H* pitch accent was the most frequently occurring pitch accent, with the H+!H* pitch accent as the second most frequent. All but one participant showed occurrences of this downstepped pitch accent. This aligns with previous work showing that the ToBI system can be applied to the speech of English-speaking children. A closer phonetic analysis of the H+!H* pitch accent resulted in the identification of three versions and two types of the H+!H* pitch accent, combining to create six phonetic sub types of this pitch accent when occurring with a low boundary tone. An initial acoustic analysis of a subset of the data lends support to the existence of these phonetic varieties.

Six phonetic varieties of the H+!H* pitch accent were presented as realizations of this pitch accent category. Importantly, although six phonetic varieties were identified, this is not a claim that these are phonologically distinct categories. Previous research shows that the H+!H* pitch accent has been difficult to tie to a single semantic interpretation, and further research is required to uncover what this pitch accent means in the speech of young children. At this point, the six phonetic realizations discovered may act as guidelines for how to label this difficult to identify pitch accent in child speech. In addition, the current analysis sheds light on how the phonetic realizations of phonological pitch accent categories are related in the case of downstepping. Downstepping has been primarily annotated by using the F0 contour as a guideline. If the multiple unique ways that F0 can occur are considered, the connection between the phonetic and phonological levels becomes more complex. Further analysis is needed to examine the semantic and pragmatic meanings behind this contour in both child and adult speech, and whether the different sub types of downstepping indicate any phonological differentiation.

5. Acknowledgements

A special thank you to the families and children who participated in the Imbrie-MIT and the Thorson-Brown University studies (Brown University study conducted under the guidance of Dr. James L. Morgan).

6. References

[1] M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language sound structure*, M. Aronoff and R. T. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 157–233.

[2] D. R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 1996.

[3] J. Pierrehumbert, "The phonetics and phonology of English intonation," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1980.

[4] K. Silverman *et al.*, "ToBI: A standard for labeling English prosody," in *Second international conference on spoken language processing*, 1992.

[5] A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, "6.9.11 Transcribing prosodic structure of spoken utterances with ToBI," presented at the MIT OpenCourseWare, Massachusetts Institute of Technology, 2006.

[6] J. Pierrehumbert and J. B. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in communication*, 1990, pp. 271–311.

[7] A. Dainora, "Modeling intonation in English: A probabilistic approach to phonological competence," *Lab. Phonol. 8*, vol. 8, p. 107, 2006.

[8] L. E. de Ruiter, "Information status marking in spontaneous vs. read speech in story-telling tasks – Evidence from intonation analysis using GToBI," *J. Phon.*, vol. 48, pp. 29–44, Jan. 2015.

[9] M. Grice, S. Baumann, and N. Jagdfeld, "Tonal association and derived nuclear accents—The case of downstepping contours in German," *Lingua*, vol. 119, no. 6, pp. 881–905, 2009.

[10] T. Rathcke and J. Harrington, "The variability of early accent peaks in Standard German," *Lab. Phonol.*, vol. 10, pp. 533–555, 2010.

[11] A. Chen and P. Fikkert, "Intonation of early two-word utterances in Dutch," in *16th International Congress of Phonetic Sciences (ICPhS 2007)*, 2007, pp. 315–320.

[12] S. Frota and M. Vigário, "The intonation of one-word and first two-word utterances in European Portuguese," in *XI International Conference for the Study of Child Language*, 2008, pp. 1–4.

[13] P. Prieto and M. del M. Vanrell, "Early intonational development in Catalan," in *Proceedings of the XVIth International Congress of Phonetic Sciences*, 2007, pp. 309–314.

[14] P. Prieto, A. Estrella, J. Thorson, and M. D. M. Vanrell, "Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish," *J. Child Lang.*, vol. 39, no. 02, pp. 221–257, Mar. 2012.

[15] J. Thorson, "The development of intonation and information structure: Perceptual and productive investigations," Doctoral dissertation, Brown University, Providence, RI, 2015.

[16] A. K. K. Imbrie, "Acoustical study of the development of stop consonants in children," Doctoral dissertation, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, 2005.

[17] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*. 2018.