AN ACOUSTIC STUDY OF WORD-MEDIAL STOP CONSONANTS IN 2-3 YEAR-OLD SPEAKERS OF AMERICAN ENGLISH

Helen M. Hanson[a] *& Stefanie Shattuck-Hufnagel*[b]

[a]ECE Department, Union College, USA; [b]Research Laboratory of Electronics, MIT, USA
hansonh@union.edu; sshuf@mit.edu

ABSTRACT

Child productions of two-syllable strong-weak words like *baby* and *cookie* are observed to exhibit a wide range of implementations of the medial stop consonant compared to adult productions. Detailed acoustic analysis of these consonants reported here lays the groundwork for prosodic analysis of the role of foot structure in the development of production of bisyllabic words. Acoustic details of the child productions differ significantly from those of the adults. The direction of these differences (e.g., longer durations, less pre-voicing) mean the child productions are more similar to an *onset* stop consonant of a two-syllable strong-weak word. However, children sometimes also *reduce* stop consonants in this environment to a greater extent than would an adult.

Keywords: word-medial, stops, phonological development, prosodic structure, binary feet

1. INTRODUCTION

When adult speakers of American English produce word-medial intervocalic stops preceding an unstressed vowel, as in *kitty* or *puppy*, the phonetic details are somewhat different from the same stop in pre-stress position. In particular, the alveolar stops /t,d/ are flapped in this position, so this context is often referred to as the flapping environment. Labial and velar stops are also reported to be affected (e.g., [5, 6]), with reduced durations overall and little to no aspiration for /p/ and /k/.

Our observation of such two-syllable words produced by the 2-3 year-olds in the Imbrie corpus [3] is that the stop consonants are produced in a wide range of different ways. At one extreme, we get the impression that the child produces the two syllables as two separate words. For example, a child production of *a baby* sounds more like an adult version of *a bay bee* than like an adult *a baby*. Figure 1 shows the waveform and spectrogram for the utterance *the puppy* (extracted from *the puppy and the dog)*, where the medial /p/ in *puppy* is produced with two separate releases and a breath is taken between them. At the other extreme, Fig. 2 shows the utterance *a puppy*, where the medial /p/ is fully voiced, with little evidence of a release.

Figure 1: An utterance of *the puppy* produced by subject C03. The medial /p/ is produced with two separate releases and an inhale between the two.

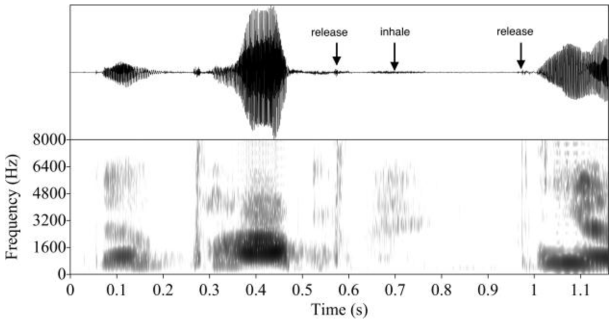
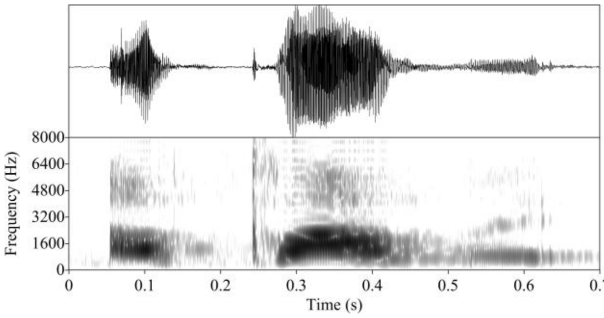


Figure 2: An utterance of *a puppy* by subject C03, with a fully voiced stop closure, resulting in an almost approximant-like production.



Examples such as these require some explanation. One possibility is that such variation is due to immature motor skills. Another is that the representation of either the word or its prosodic structure is incorrect or in flux. Or, the representation could be correct, but the phonological-to-phonetic processes that govern its production are still being learned.

Turk's [5] results indicate that flapping is foot-conditioned: alveolar stops can be flapped only when they occur in non-foot-initial position [4]. This at least raises the possibility that the reduction seen in the labial and velar stops may also be foot-conditioned. Therefore, we wonder if the acoustic extremes that we have observed in the two-syllable word-medial stops of children may be related to their immature prosodic structural representations. One possibility is that they do not yet recognize the foot as a prosodic structure. However, the literature on the prosodic structure of developing speech indicates that while children's early grammars may not contain all possible prosodic structures, the binary foot is represented early on. Evidence for this includes the observation that young children can more reliably produce an article when it can form a binary S-W foot with the preceding word (e.g., *sing a song* v. *singing a song*, [1, 2]). Nonetheless, the fact that a child is able to represent binary feet and use them to plan and structure speech utterances does not necessarily mean that the child can produce all of the appropriate acoustic cues to foot structure that an adult might use.

This paper is a first step toward explaining our observations on developing medial stop consonants. It lays the groundwork for subsequent study of the prosody of child productions of these two-syllable words by examining in detail the acoustic characteristics of word-medial consonants. Our immediate goal is to determine ways in which child productions differ from adult productions.

2. EXPERIMENTAL METHODS

2.1. Corpus

Data are drawn from the Imbrie Corpus [3]. Recordings were made from 10 children and the primary female caregiver of each child. The children were recorded approximately monthly over a six-month period, while the caregivers were recorded just once. The age of the children at the first recording ranged from 2;6 to 3;3. (Further details about the subjects and the recordings can be found in [3], available online.) The twenty target words were elicited multiple times from each child during play sessions in which the experimenter prompted the child using pictures and objects.

Each token of a target word was excised from the play session recording and categorized by whether the token was elicited in a naming task or in a longer statement. The total number of tokens of each target word per session varied widely from zero to more than 20.

The words used in this study are *baby*, *daddy*, *Maggie*, *puppy*, *kitty*, and *cookie*. Only tokens recorded during the first and last play sessions are analyzed here. Tokens with excessive background noise were discarded. Data for four child participants (and their caregivers) are presented in this paper.

2.2. Acoustic labeling and measures

The Praat software was used to display the data and to label relevant speech events, including vowel onset and offset, offset of pre-voicing, closure and releases of both onset and word-medial stops, and the onset and offset of aspiration noise. The labels were saved in TextGrid files, which were converted to records in a MATLAB database. Several acoustic measures were then derived and stored in the database: V1 duration, V2 duration, V1 duration/V2 duration, medial stop closure duration, medial stop VOT, medial stop aspiration, and % of medial stop that is voiced.

2.3. Analyses

Our first analysis was to determine if measures on the child data changed during the six months between the two recordings. A repeated measures ANOVA with Speaker as the random factor and Voicing, Place, Play session, and Context (naming v. statement) showed that Play session was a main effect for several measures. Therefore, we could not pool the child data over play sessions. Instead, two more analyses were done to compare (1) the child data from the first play session with the adult data, and (2) the child data from the last play session with the adult data. Both analyses were repeated measures ANOVAs with Family as the random factor and Voicing, Place, Context, and Age (adult v. child) as fixed factors.

3. RESULTS

The results of the first ANOVA are summarized in Table 1, where *p* values are given for the factors that had significant effects on the acoustic measures. As expected, both underlying Voicing and Place of articulation have significant effects on the closure, VOT, aspiration, and % voicing of the medial stop consonant. Voicing is also seen to have an effect on the duration of V1. Of greatest interest to us is the fact that Age group has an effect on all acoustic measures except for the medial VOT. Therefore, in play session 1, the acoustic details of the child productions differ from those of the adults. Table 2 lists the *p* values for the interactions between Voicing, Place, and Age. For every measure except VOT, Age interacts with Voicing or Place, indicating that their effects on the acoustic measures differ for adults and children.

Table 1: Main effects of ANOVA comparing adult data with child data from the first play session. A blank cell indicates an insignificant effect. The factors of Family and Context had no significant effects for any measures and thus are omitted.

	Voicing	Place	Age
V1 dur	p<0.001		p<0.001
V2 dur			p<0.05
V1/V2	p<0.001		p<0.01
Medial Closure	p<0.001	p<0.001	p<0.001
Medial VOT	p<0.001	p<0.001	
Asp. Dur.	p<0.001	p<0.001	p<0.05
% voiced	p<0.001	p<0.001	p<0.01

Table 2: Interactions of the fixed factors for the comparison of adult data with child data from the first play session. A blank cell indicates an insignificant effect. The factor Context did not interact with any other factors and thus those interactions are omitted.

	Voicing* Place	Voicing *Age	Place* Age
V1 dur	p<0.01	p<0.001	
V2 dur	p<0.05	p<0.001	
V1/V2			p<0.05
Medial Closure	p<0.001	p<0.05	p<0.001
Medial VOT			
Asp. Dur.			p<0.001
% voiced	p<0.001	p<0.05	p<0.05

Figure 3 illustrates the effects of Voicing and Age on the duration of V1. For both Age groups, V1 is longer before voiced medial consonants then before unvoiced. (The difference could in part be due to the fact that the vowels preceding the voiced stops are tense, while the vowels preceding the unvoiced stops are lax.) V1 is also longer for children than for adults. In addition, Fig. 3 illustrates the interactions between factors. The difference between the two voicing conditions is greater for the children than for adults. The interaction of Voicing and Place appears to be due to a shorter V1 duration before /b/ than before /d, g/ for the adult speakers; this difference is most likely due to the fact that the vowel preceding /b/ is /e/, while the vowel preceding /d, b/ is /æ/.

It is clear from Fig. 3 that the children correctly use V1 duration to cue the voicing of the medial consonant. But it is equally clear that they use a more exaggerated form of the cue than the adults. One could argue that the children may simply be speaking more slowly than the adults, but if that were the case, the ratio V1/V2 should be about the same for both adults and children. However, as Table 1 shows, it is not the case.

Figure 3: Duration of the vowel in the first syllable of a two-syllable strong-weak word. Error bars indicate standard error.

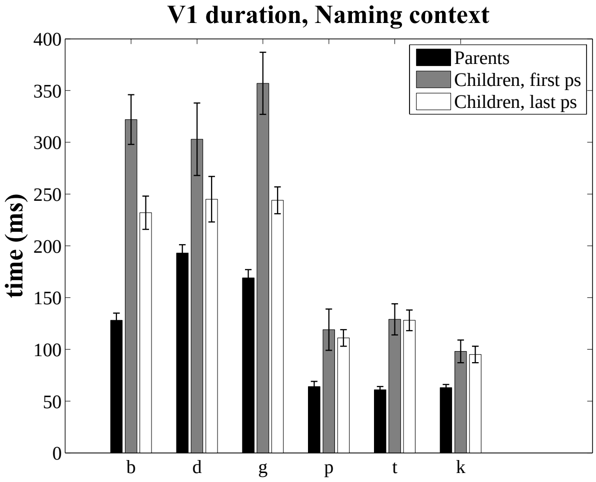


Figure 4 illustrates the effect of Voicing, Place, and Age on the percentage of the medial closure that is voiced (percentages that are greater than 100 simply indicate that voicing extends beyond the release into the VOT). Both age groups have more pre-voicing for the voiced stops and for /t/, the latter suggesting that children are consistently applying the flapping rule in the two-syllable strong-weak environment. However, the interactions of Voice, Place, and Age indicate that the children are not yet producing adult-like stops in this environment. For example, the pre-voicing for /d, g, t/ is not as extensive for the children as for the adults.

The results of the second ANOVA are summarized in Tables 3-4, where *p* values are given for the factors that had significant effects or interactions on the acoustic measures. The effects of Voicing and Place are nearly the same as they were in the first comparison (Tables 1-2), as could be expected. However, Age no longer has an effect for two of the measures, the duration of medial aspiration and % of medial closure that is voiced, indicating that some aspects of the child productions have changed in ways that make them more similar to those of the adults. For example, Fig. 3 shows that although V1 duration is still significantly different for adult and child voiced stops, the child V1 durations are reduced compared to the first play session. Likewise, Fig. 4 shows that the percent of voicing of the medial closure is greatly increased for the child productions of /b, d, g/, to the point that they are well within the range of the adults. Interestingly, the percent of voicing for /p, k/ have also increased, away from the adult data, which might indicate that some children are now *over* reducing their stop productions in this environment (see Fig. 2 for an example of a medial /p/ with excessive prevoicing).

Figure 4: Percent of medial closure that is voiced. Error bars indicate standard error.

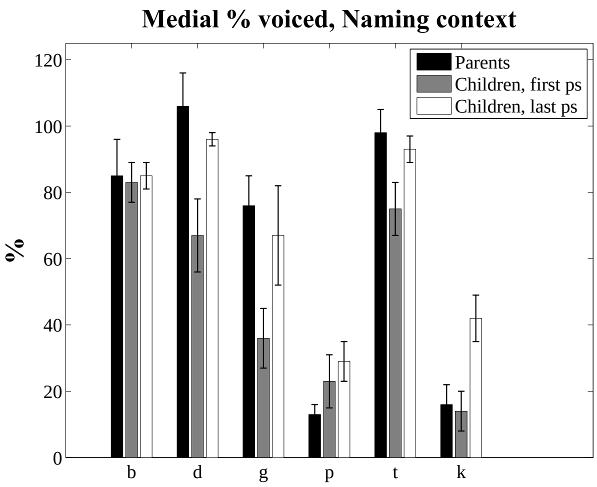


Table 3: Main effects of ANOVA comparing adult data with child data from the last play session. A blank cell indicates an insignificant effect. The factor of family was not significant for any measures and thus is omitted. Context had only weak effects on a few measures, and in the interest of space it is also omitted.

	Voicing	Place	Age
V1 dur	p<0.001		p<0.001
V2 dur			p<0.001
V1/V2	p<0.001	p<0.05	p<0.001
Medial Closure	p<0.05	p<0.001	p<0.001
Medial VOT	p<0.001	p<0.001	
Asp. Dur.	p<0.001	p<0.001	
% voiced	p<0.001	p<0.001	

4. DISCUSSION AND FUTURE WORK

Although space limitations prevent us from discussing all aspects of the results in detail, it is clear from Tables 1-4 and Figs. 3-4 that the acoustic details of the child subjects differ from those of their caregivers. Moreover, the direction of these differences (longer durations, less pre-voicing) make the child productions more similar to an *onset* stop consonant of a two-syllable strong-weak word, which may be why the child productions sound more like two strong syllables. However, we know from examples such as the utterance shown in Fig. 2 that children sometimes *reduce* stop consonants in the flapping environment to a greater extent than would an adult; evidence for that can also be seen in Fig. 4. In future work, we will analyze data for additional subjects and play sessions in order to get a more detailed picture of the nature and development of stop consonants in this environment.

Table 4: Interactions of the fixed factors for the comparison of adult data with child data from the last play session. A blank cell indicates an insignificant effect. The factor context had weak interactions with the other factors for some measures, but these are omitted due to space limitations.

	Voicing* Place	Voicing *Age	Place* Age
V1 duration	p<0.05	p<0.001	
V2 duration			
V1/V2			p<0.05
Medial Closure	p<0.01	p<0.05	
Medial VOT	p<0.001	p<0.05	
Asp. Dur.	p<0.001	p<0.001	p<0.01
% voiced	p<0.001		p<0.05

5. REFERENCES

[1] Demuth, K. 1996. The prosodic structure of early words. In Morgan, J.L., Demuth, K. (eds.), *Signal to Syntax*. Mahwah, NJ: Lawrence Erlbaum, 171-184.

[2] Gerken, L. 1996. Prosodic structure in young children's language production. *Language* 72, 683-712.

[3] Imbrie, A. 2005. *Acoustical Study of the Development of Stop Consonants in Children*. Ph.D. thesis, MIT, Cambridge, MA. (*http://hdl.handle.net/1721.1/27198*)

[4] Kiparsky, P. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10, 421-441.

[5] Turk, A. 1992. The American English Flapping Rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetic Laboratory* 7, 103-133.

[6] Warner, N., Tucker, B.V. 2007. *Categorical and Gradient Variability in Intervocalic Stops.* Presented at the *Linguistic Soc. of America Annual Meeting* Anaheim, CA.