BUCLD 39 Proceedings
To be published in 2015 by Cascadilla Press
Rights forms signed by all authors

18 Month-Olds Compensate for a Phonological Alternation

Adam J. Chong and Megha Sundara*

1. Introduction

The input to a child contains regular variant productions of a word that while not canonical are systematic and predictable from context. Allophonic variation is one such regularity in the input. In this paper, we investigate toddlers' abilities to compensate for allophonic variation. There is some evidence that at least after they have started producing connected speech, toddlers are able to compensate for phonological alternations in their native language. Skoruppa, Mani, Plunkett, Cabrol & Peperkamp (2013) tested 24-month-old English and French-learning children's sensitivity to one such regularity in the input motivated by ease of articulation: assimilation (in which phonological features of one sound spread onto neighboring sounds). They used a variant of the intermodal preferential looking task (IPL: Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987) to examine 24-month-olds' sensitivity to assimilations and mispronunciations in sentence-medial position. In this study, French toddlers were able to recover the correct target word when presented with a voicing assimilated form in a licensing context (e.g. French: *Regarde le buz* ***d****evant toi,* 'Look for the bus in front of you'). In contrast, English toddlers when presented with a similar type of voicing assimilation (e.g. English: *Can you find the shee***b th***ere?*) failed to recover the target item.

American English presents another example of phonological variation: tapping (or flapping). Word-final /t/s and /d/s in American English can be produced as a tap [ɾ] in a number of environments. This most commonly occurs when the following morpheme or word begins with an unstressed vowel, as in *-ing*, *again* or *in* (Kahn, 1980; Oshika, Zue, Weeks, Neu & Aurbach 1975; Turk, 1992; Zue & Laferriere, 1979). In a recent study, Sundara, Kim, White

---

* Adam J. Chong and Megha Sundara, Department of Linguistics, University of California, Los Angeles. Address all correspondence to AJC at ajchong@ucla.edu.
Presented at BUCLD 39 as 18-month-olds compensate for a phonological alternation. We are grateful to Anya Mancillas, Robyn Orfitelli and RAs of the UCLA Language Lab for their help in testing participants. We also thank members of the UCLA Phonetics Lab, particularly Pat Keating, Bruce Hayes and Yun Jung Kim as well as audiences at the LSA Workshop on Variation in the Acquisition of Sound Systems and the Symposium on Cognitive and Language Development 2014 for feedback. This work was supported by NSF BCS-0951639 and UCLA COR Faculty Research grants awarded to MS.

and Chong (2013) examined whether or not 12-month-old infants were able to successfully map tap forms to both /t/ and /d/ forms (i.e. recover *pat* or *pad* from surface *patting* or *padding*). Using a word segmentation paradigm (Jusczyk & Aslin 1995), they found that when 12-month-olds were familiarized with *padding*, they listened significantly longer to *pad* but not *pat*. Moreover, infants successfully discriminated between [d] and [ɾ]. 12-month-old infants, thus, were already able to equate both [ɾ] and [d] forms of words, but not [ɾ] and [t] forms.

Sundara et al.'s (2013) results suggest that at least for [ɾ] variants in English, the ability to undo regular variation in perception might be in place at an earlier age than 24-months. In this paper, in view of these findings, we investigate younger children's lexical representations to examine the extent of their knowledge of variation before they produce connected speech. Evidence from a number of studies has shown that between 14- and 18-months, toddlers are already sensitive to mispronunciations in both the onset and coda of words (Bailey & Plunkett, 2002; Swingley & Aslin, 2000; Swingley, 2009) and also the *severity* of deviance from the correct pronunciation of a known word (Ren & Morgan, 2011; White & Morgan, 2008). Moreover, toddlers at this age also show some ability to factor out phonetic variation that does not alter the meaning of a word across different unfamiliar accents (Best, Tyler, Gooding, Orlando & Quann, 2009; Best, Tyler, Kitamura & Bundgaard-Nielsen, 2010; Mulak, Best, Tyler, & Kitamura, 2013). Therefore it is possible that the ability to recognize familiar words produced with a regular phonological variant is in place some time around the same age.

In this study, we assessed 18-month-olds' sensitivity to alternating segments in coda position, specifically the alternation between [t], [d] and [ɾ]. We followed the IPL experimental procedure set out in White & Morgan (2008). Toddlers were shown pairs of pictures on a screen, one familiar (e.g. a cat) and the other unfamiliar (e.g. dumbbells), and during the test phase the familiar object was labeled. In this paradigm, toddlers looked longer at the familiar object when presented with a correct pronunciation of its label than a mispronunciation. Moreover, toddlers showed graded sensitivity effects to the size of mispronunciations. Using this paradigm should allow us to see if there is an effect of regularly alternating segments on word recognition. Specifically, we examined whether or not toddlers are able to recover the correct target word when the target word is produced with a [ɾ] as opposed to canonical [t] or [d].

Importantly, because the alternation between [t], [d] and [ɾ] occurs across a word boundary, toddlers had to spontaneously detect mispronunciations and alternations in sentence-medial contexts. Sentence-medial context has been known to be difficult for toddlers (Plunkett, 2005). Because sentence-medial contexts are difficult for children to process, Skoruppa et al. (2013) exposed toddlers to the familiar words in a pre-test training phase before the words were presented in sentence-medial position. In our present study, toddlers were not exposed to the familiar words ahead of time. To establish a baseline from which to compare children's responses, we tested adults using the same stimulus materials. Based on previous findings (Pitt, Dilley & Tat, 2011; Ranbom, Connine & Yudman, 2009), we expected adults to recognize the intended target word regardless of whether or not it was presented with a canonical [d], [t] or a [ɾ] in the appropriate context as compared to a true mispronunciation.

2. Experiment 1: Adults

38 adult participants were recruited via the UCLA Psychology Subject Pool. Subjects received course credit for participation. All identified themselves as native speakers of American English. Subjects were randomly assigned into one of 4 counterbalancing groups.

2.1 Stimuli

Since our final goal was to test children using the same paradigm, we chose target words that were familiar to 18-month-old toddlers. The target words were a set of 6 /t/-final and 6 /d/-final nouns that were reported to be familiar to toddlers at 18-months of age, based on the lexical norms of the MacArthur Bates Communicative Index (CDI; Dale & Fenson, 1996). A list of target words is shown in Table 1 with their CDI comprehension frequencies at 16-months (the oldest age for which these norms are available) and CDI production frequency at 18-months respectively.

Table 1. Comprehension frequency of target [t]- and [d]-words for 16-month-olds and production frequency for 18-month-olds

Target word		CDI comprehension freq. 16 months (%)	CDI production freq. 18 months (%)
		[t]-words	
1	hat	53.7	61.1
2	cat	51.2	76.4
3	boat	50	N/A
4	foot	41.2	50
5	bat	17.5	N/A
6	plate	13.8	33.3
		[d]-words	
1	bird	67.5	79.2
2	bed	28.8	68.1
3	bread	25	48.6
4	food	16.3	47.2
5	cloud	12.5	N/A
6	slide	12.5	48.6

Unfamiliar items were similarly nouns which were real world objects similar in category status (e.g. objects, animals) and visual complexity to the familiar nouns. The names of the unfamiliar items were not known by toddlers at this age, according to the MacArthur CDI (Dale & Fenson, 1996).

Additionally, we included 6 nonsense words that were phonotactically legal in English but phonologically dissimilar to the names of the items they labeled: *tesh* [tɛʃ], *wiss* [wɪs], *bize* [baɪz], *dape* [deɪp] (from Albright & Hayes, 2003), *lif* [lɪf] and *neem* [nim] (from Stager & Werker, 1997).

The audio stimuli were digitally recorded by a phonetically-trained female native speaker of American English in an infant-directed register. Target words were recorded in two sentence frames: (1) *Look for the [target] Sam lost!* and (2) *Look for the [target] again!* Sentence frame (1) provided the appropriate pre-consonantal context for the [t] or [d] variant, whereas (2) prompted the appropriate context for tapping which usually occurs when /t/ follows a vowel and precedes an unstressed vowel (e.g. Kahn 1980; Turk 1992). Furthermore, we made sure that there was no prosodic break following the target word in the tap condition since this has been known to disrupt the occurrence of taps (de Jong, 1998). Following Herd, Jongman & Sereno (2010), we ensured that that the closure durations for stop and tap variants formed a bimodal distribution with at least one time-bin of 10 ms separating them (Fig. 1).

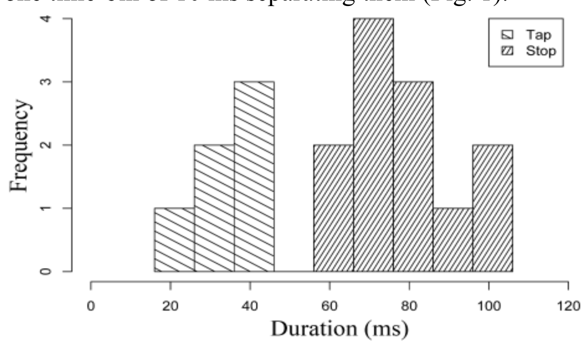

Fig. 1. Closure duration of stop and tap stimuli

Mispronunciations of target words were always one-feature place mispronunciations. These productions always contained the same vowel as in the stop and tap conditions. The mispronounced target words were also recorded in both sentence frames. A list mispronunciations and their corresponding correct pronunciation is shown in Table 2 below. In addition to the target sentences, two more sentences were recorded, *Do you see it?* and *Can you find it?* These sentences played following the offset of the target sentences to maintain participants' interest on the objects on the screen. To make the task more natural, we played a video of a puppet monkey, *Sam*, and recorded a short passage to introduce him. All sentences were recorded at 44,100 Hz were scaled to 65 dB in Praat (Boersma & Weenink, 2011) and played at a comfortable listening level of 65 dB.

Finally, visual stimuli were digitized color images, presented side by side on a 21.5 inch display monitor. Images were 350 by 350 pixels in size. Each familiar image was yoked with a unique image that was unfamiliar to toddlers in every trial. An example pair of images is shown in Fig. 2.

Table 2. One-feature mispronunciation of test items

/t/-words				/d/-words			
Correct Pron.		Mispronunciation		Correct Pron.		Mispronunciation	
cat	[kæt]	*cak*	[kæk]	slide	[slaɪd]	*slige*	[slaɪg]
bat	[bæt]	*bap*	[bæp]	bed	[bɛd]	*beb*	[bɛb]
foot	[fʊt]	*foop*	[fʊp]	bread	[bɹɛd]	*breag*	[bɹɛg]
plate	[plet]	*plake*	[plek]	food	[fud]	*foog*	[fug]
hat	[hæt]	*hap*	[hæp]	cloud	[klaʊd]	*cloub*	[klaʊb]
boat	[boʊt]	*boak*	[boʊk]	bird	[bɝd]	*birb*	[bɝb]

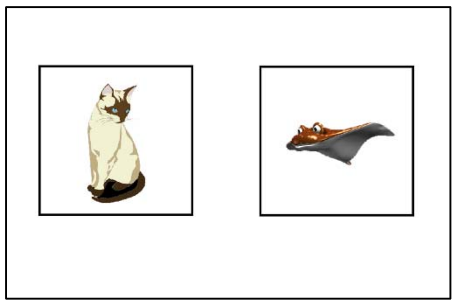

Fig. 2. Sample pair of display images (familiar target = *cat*, distractor = stingray). Boxes represent interest areas.

2.2 Procedure

We followed the experimental procedure set out in White and Morgan (2008). Participants were seated on a chair facing a 21.5 inch Asus display monitor with a 2-ms refresh rate and an SR Eyelink 1000 (SR Research, Mississauga, Canada). They were seated between 500 and 600 mm away from the eyetracking camera (*M = 553, SD =16.81*). Speech stimuli were presented through concealed a set of Cyber Acoustics CA-3602 loudspeakers hidden behind curtains under the display in front of the participants. Participants' fixations were recorded by the Eyelink system, using the arm-mount remote configuration. At the beginning of each experimental session, participants' gaze was calibrated and validated using a five-point calibration. Validations of Fair and Good as determined by the eyetracking software were accepted.

After calibration and validation, participants saw a short video with a puppet monkey, Sam. Participants were told that Sam had lost his toys and they were helping him find them. Each trial consisted of two phases: (1) a silent baseline and (2) a test phase. Participants were initially centered with a looming object on the screen. Once participants fixated to the center of the screen, the experimenter initiated the baseline phase. During the silent baseline phase, participants saw two images horizontally side-by-side 960 pixels apart. Images were always pairs of a "familiar" and "unfamiliar" image for an 18-month-old and appeared simultaneously on the screen. They remained on the screen for 4 seconds. The baseline phase was included to allow subjects time to look at both objects and to allow us to gauge any baseline preferences for a particular image.

Following this, participants were recentered with the same looming object and once they fixated to the center of the screen, the test phase started. Participants were shown the same pair of images they saw in the baseline. At the onset of image presentation, participants heard the target sentences. They then heard one of the follow-up sentences, *Do you see it/Can you find it?,* 750 ms after the offset of the target sentence. Images remained on the screen for another 2 seconds. Hence, the test phase lasted approximately 6-seconds.

The experiment consisted of 24 trials in total. 6 trials were in the novel condition. These were included to encourage participants to look at the distractor item and remained constant across subjects. The remaining 18 trials were divided evenly between the three crucial test conditions: stop, tap and mispronunciation, with 6 trials in each condition. In order to control for repetition effects, no participant saw the same target word in both the canonical stop and tap conditions. 4 counterbalancing groups were created with the assignment of stimulus pairs to each condition counterbalanced across subjects. The 6 /t/ and 6 /d/ target words were divided into two groups of 3 words each. The targets in the MP condition were d always the same as in the tap condition.

The order of trials was randomized using Experiment Builder (SR Research, Mississuaga, Canada). Side of presentation was also randomized such that target images appeared equally on both the left and right side of the screen. The location of target images was kept consistent across phases. Each experimental session lasted about 4 minutes.

2.3 Results and discussion

Two interest areas (Target and Distractor) were set at 500 by 500 pixels around the target image (see Fig. 2). The interest areas were larger than the images themselves to ensure that fixations around the images, while not on the images themselves would be included. Fixations to the target were extracted for the entire 4-s duration of the baseline phase. For the test phase, looking behavior was coded for the 2-s following either the onset or coda consonant depending on the analysis. This was the approximate time window used by Swingley (2009). In order to facilitate comparison across adults and toddlers, the same time-window was used in Experiments 1 and 2. Trials were eliminated if participants did not look at both subjects during the baseline phase as defined by fixation proportions less than 0.1 or more than 0.9 (14 trials, 1.9%). Participants were excluded entirely if more than 50% of their trials were not useable due to the above criteria (3 of 38). Another three participants were excluded due to experimenter error. This yielded a final pool of 32 participants divided evenly across the 4 counterbalancing groups, with 8 participants in each group.

The dependent measure was the difference in participants' looking to the familiar object between the silent baseline phase and the test phase (%Test - %Baseline). This measure indicates therefore how much subjects looked to the target object after it was named, taking into account any baseline preferences for that object. A difference score was calculated for each trial. The difference scores were then subject to analyses using linear mixed-effect models with the *lme4* package (Bates, Maechler & Dai 2008) in R (R Development Core Team, 2008). The anova() function was used to compare models which were in a subset relationship. The results of this likelihood ratio are reported as chi-squares.

In the first analysis (Fig. 3), we were interested in comparing adult listeners' behavior in the stop and novel conditions as a means of validating the task. Here we compared looking behavior starting from the start of the first consonant of the target word as the target and novel word differed in their initial segment. Test-baseline difference scores are shown in Fig. 3. As expected, adult listeners looked significantly more to the target object in the test phase compared to baseline in the canonical stop condition than in the novel condition ($\chi^2(1) = 51.47$, $p < 0.001$). In fact, when adult listeners heard a novel label for the familiar target on the screen, they looked more at the distractor object. This result validates the basic paradigm.

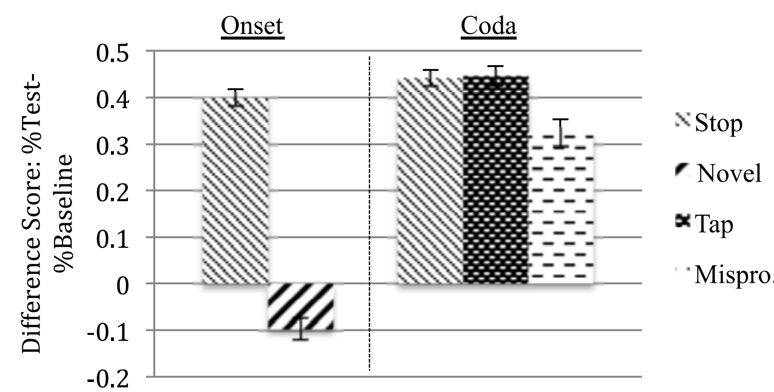

Fig. 3. Experiment 1 difference scores: (L) onset analysis; (R) coda analysis

In the second analysis, we compared participants' behavior when they heard target words produced with a stop, a tap and a one-feature mispronunciation (Figure 3). Looking behavior was analyzed from a 2-s window starting from the start of the coda consonant (following the offset of the vowel). Recall that the stop, tap and mispronunciation trials differ in their coda consonants. From Figure 3 we can see that listeners looked more to the target image in the test compared to baseline in *all* three conditions, suggesting that they were all treated as potential labels for the target objects.

We included two factors in our model, Word type (/t/ vs. /d/) and Condition (Stop vs. Tap vs. Mispronunciation), as well as, their interaction. Condition was sum-coded with the Tap condition as reference. The interaction of Word Type by Condition was not significant ($\chi^2(2) = 3.35$, $p = 0.19$). This interaction term was excluded in subsequent model comparisons. Test of the main effects of Word Type and Condition were conducted against a superset model with both variables but without the interaction term. Participants' looking behavior did not differ by Word Type ($\chi^2(1) = 1.18$, $p = 0.28$), indicating that they treated /t/ and /d/ words similarly across conditions. Importantly, there was a significant effect of Condition ($\chi^2(2) = 14.54$, $p < 0.001$).

Post-hoc pairwise comparisons were conducted using the *glht()* function of the *multcomp* package (Hothorn, Bretz & Westfall 2008). All comparisons were adjusted using Tukey's method. The analysis revealed that listeners looked significantly more at the target object in the stop condition and tap condition than in the mispronunciation condition (Stop vs. Mispronunciation: $p < 0.001$; Tap vs. Mispronunciation: $p < 0.001$). Importantly, listeners did not differ in behavior in the stop and tap conditions ($p = 0.99$), looking equally at the target image at test.

Adult American English listeners therefore recognize words equally well when these words are produced with either a canonical stop variant or tap variant. Furthermore, mispronounced words incur a cost to word recognition when compared to *both* the stop and tap conditions although they are nonetheless accepted as labels for the target image. These results provide a baseline from which to compare the behavior of the 18-month-old toddlers in Experiment 2.

3. Experiment 2: 18-month-old toddlers

38 English-learning toddlers (mean age: 18.2; range: 17.6: 20.2 months) were tested in this study. An additional 27 toddlers were tested but were excluded for fussing out and moving too much to secure a good track (n=16), wrong language background (n=5) and not having enough critical test trials (n=7).

3.1 Stimuli

The same target words used in Experiment 1 were used in Experiment 2. Initial piloting revealed that toddlers found the context used for canonical stop productions, *Look for the X Sam lost*, difficult to process. In order to make the task easier, we used the temporal adverb *now* instead, resulting in the following sentence frame: *Look for the X now!* Audio stimuli in this condition were re-recorded by the same female speaker used in Experiment 1. Similarly, the same visual stimuli were used in this experiment as in Experiment 1.

3.2 Procedure

The procedure was largely the same as in Experiment 1. Toddlers were seated between 500 and 600 mm away from the eyetracking camera ($M = 560$, $SD = 45$). Following the experiment, parents were asked to complete a language background questionnaire, a production vocabulary form (Dale & Fenson 1996) and a questionnaire that rated the familiarity of the images on the screen.

Caregivers were asked to rate pictures on a scale of 1 to 5 with a score of 1 indicating that both the picture and word are *not familiar* to the child and 5 indicating that both the picture and word are *highly familiar* to the child.

Initial piloting found that toddlers had trouble sitting through the entire experiment including the time it took for experimental set-up. Thus experiment 2 consisted of only 22 trials, with only 4 novel trials instead of 6 (*tesh* and *wiss* were excluded). The number of trials in the test conditions was kept the same.

3.3 Results and discussion

Results from the parental questionnaire indicated that, on the whole, our chosen familiar images were indeed familiar to the toddlers in our study. The average ratings of familiar images were 4.1 (SD = 0.66) and those of unfamiliar images were 1.2 (SD = 0.26). Although familiar images were indeed familiar to the toddlers in our study, a couple of images had lower scores than other images (e.g. *bat* = 2.3, *cloud* = 2.9). Trials were only included if parents rated those images with a 3 or above (i.e. that the word, but not the images was familiar to the child; 14% of trials). Furthermore, we excluded trials in which toddlers did not look at either image enough in the baseline (19% of trials). Finally, trials were excluded if the toddler moved significantly in the test phase resulting in a lost of the tracking of the eye (7% of trials). Eight toddlers were excluded entirely since less than 50% of their trials were usable following the above exclusions. A total of 601 trials from 38 toddlers were included in the final analysis.

Test-baseline differences scores were calculated for each trial in the same way as in Experiment 1 (Fig. 4). A comparison of toddlers' fixation behavior in the canonical stop vs. phonologically dissimilar novel condition showed that toddlers looked significantly more to the familiar image in the canonical stop condition than they did in the novel condition ($\chi^2(1) = 5.95$, $p = 0.015$). As with adults, toddlers looked more at the unfamiliar distractor image when they heard a novel label, validating the general paradigm used.

The crucial comparison was between the stop, tap and mispronunciation conditions. As in Experiment 1, our model included Word Type and Condition as factors. Since there was no interaction between Word Type and Condition ($\chi^2(2) = 0.36$, $p = 0.83$), this factor was taken out in subsequent model comparisons. Test of the main effects of Word Type and Condition were conducted against a superset model with both variables but without the interaction term. There was a significant effect of Word Type with toddlers looking more at the familiar image when they heard /d/ words overall than /t/ words ($\chi^2(1) = 4.11$, $p = 0.046$). The effect of condition was only marginally significant ($\chi^2(1) = 5.353$, $p = 0.069$). Post-hoc comparisons using the *glht()* function of the *multcomp* package (Horton et al., 2008) revealed that there was no significant difference between fixations in the stop and tap condition (p = 0.93). While toddlers looked more at the familiar image in both the stop and tap conditions than in the mispronunciation condition, these differences did not reach significance (stop: $p = 0.18$; tap: $p = 0.08$).

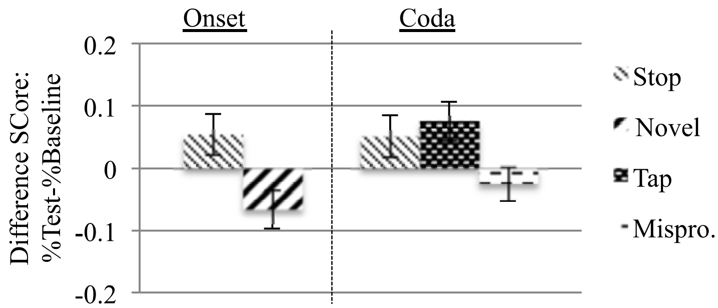

Fig. 4. Experiment 2 difference scores: (L) onset analysis; (R) coda analysis

Toddlers' poor performance on the stop condition was surprising given that they showed significant differences when compared to the novel condition. It is possible that this poorer performance is due to the fact that that target words are embedded in sentence-medial position, in some cases, pre-consonantally. Skoruppa et al. (2013) faced similar difficulty in an initial pilot study where they found no significant differences between correctly pronounced and mispronounced lexical items in a similar task. In that study, toddlers were familiarized with the target lexical items in the experiment to aid toddlers in the task. Toddlers in our current study did not receive any familiarization with the target items.

A further factor that might have influenced toddlers' performance in this task could have been due to differences in expressive vocabulary as ascertained through the MCDI (Dale & Fenson, 1996). Toddlers' vocabulary size has previously been linked to their ability to recognize variant forms of words in spoken word recognition tasks (e.g. Best et al., 2010; Mulak et al., 2013; Swingley, 2009; Werker, Fennell, Corcoran & Stager, 2002). It is possible then that toddlers' vocabulary size in our current study may have impacted their performance on their task.

MCDI expressive vocabulary scores were collected after each experimental session. The average raw production score was 24.3. We performed a median split (median = 20) of the entire subject population with 19 toddlers in both the high vocabulary and the low vocabulary groups. A separate analysis was conducted for each group utilizing the same procedure as above.

In the low vocabulary group (Fig. 5), there was neither a significant effect of Word Type ($\chi^2(1) = 0.18$, $p = 0.67$) nor of Condition ($\chi^2(2) = 1.34$, $p = 0.51$). There was similarly no significant Word Type X Condition interaction ($\chi^2(2) = 0.31$, $p = 0.86$). In the high vocabulary group, however, there was a significant effect of both Word Type ($\chi^2(1) = 5.75$, $p = 0.017$) and Condition ($\chi^2(2) = 7.38$, $p = 0.025$). There was, however, no significant interaction ($\chi^2(2) = 1.33$, $p = 0.515$). Post-hoc comparisons of the effect of condition showed that high-vocabulary toddlers looked significantly more at the familiar image when they heard a stop form than when they heard mispronounced word form ($p = 0.03$). Although they looked numerically more at the familiar image when they heard a tap form than when they heard a mispronounced form, this effect was only marginally significant ($p = 0.07$). Importantly, looks to the familiar image did not differ between the stop and tap conditions ($p = 0.93$).

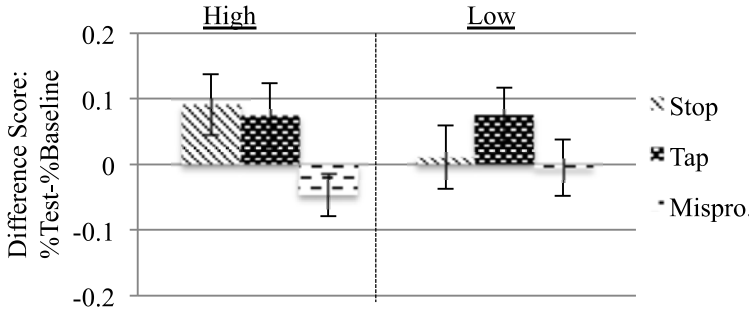

Fig. 5. Difference scores by CDI: (L) High vocabulary; (R) Low vocabulary

Thus, the results of this experiment provide evidence that 18-month-old toddlers are able to recognize familiar words produced as a regular tap variant, as evidenced by the equivalent looking times in both the tap and stop conditions. Yet the ability to accept this phonological variant seems to not be entirely in place with some toddlers continuing to treat these forms as mispronunciations. Moreover, the ability to perform the task on the whole seems to be impacted on by toddlers' vocabulary size.

4. General Discussion

Using an intermodal preferential looking task, we investigated whether adults and 18-month-old English-learning toddlers are able to recognize a familiar word produced with a regular phonological variant, namely a tap. American English adult listeners recognized a target word equally well whether the word was produced with a stop or tap variant. Similarly, 18-month-old toddlers recognized target words equally well with both the stop and tap variants, although this ability was moderated by vocabulary size, with toddlers with a higher vocabulary being better at the task on the whole.

Our results add to the growing body of experimental results indicating that by around 18-months-old, toddlers have detailed phonetic representations such that they reject words with a minimally mispronounced coda consonant (e.g. Swingley, 2009). Moreover, we provide further evidence for burgeoning *phonological constancy* across a number of pronunciation variants at this age (Best et al. 2009, 2010; Mulak et al., 2013), which seems to interact with individual children's expressive vocabulary. Our results also give us an indication of the time frame within which the [t]~[ɾ] alternation is learned.

Recall that Sundara et al. (2013) found that 12-month-olds do not show an ability to map [ɾ] onto /t/. 18-month-olds in our study, however, were able to successfully recognize /t/-words produced with a [ɾ]. This therefore gives an indication that the ability to recognize these forms develops some time in the first half of the second year of life.

The current study, however, is limited in its ability to pinpoint the exact mechanism that is used to learn that tap forms are regular phonological variants of /t/. As Sundara et al. (2013) showed, 12-month-olds succeeded at mapping tap forms to /d/ but not [t]. Specifically, they argued that this resulted from a bias to treat taps as /d/s due to their perceptual similarity. An open question then is how American English children come to overcome this bias and eventually learn that taps map to /t/ as well. Is their learning item-specific and dependent on their knowledge of known words? That is, is it necessary for toddlers to have heard individual words produced with tap forms, or do they learn a general rule and are thus able to extend this to novel words? Furthermore, it is unclear how important context is in learning of these variants. The current study only presented tap variants in a context which supports it. It is possible, however, that children encode the entire word form in learning without necessarily taking the surrounding phonological context into account. Future work will seek to clarify these questions.

References

Albright, Adam, and Bruce Hayes (2003). Rules vs. Analogy in English Past Tenses: A computational/experimental study. *Cognition, 90,* 119-161.

Bailey, Todd M., and Kim Plunkett (2002). Phonological specificity in early words. *Cognitive Development, 17,* 1265-1282.

Bates, Douglas, Martin Maechler and Bin Dai (2008). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. <http://lme4.rforge.r-project.org/>

.Best, Catherine T., Michael D. Tyler, Tiffany N. Gooding, Corey B. Orlando and Chelsea A. Quann (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science,* 20, 539–542.

Best, Catherine T., Michael D. Tyler, Christine Kitamura and Rikke L. Bundgaard-Nielsen (2010). *Vocabulary size at 17 months and the emergence of phonological constancy in word recognition across native and nonnative dialects.* Presented at the International Conference on Infant Studies, Baltimore, MD.

Boersma, Paul, and David Weenink (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.45, retrieved 21 May 2011 from http://www.praat.org/

Dale, Philip S., and Larry Fenson (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125-127.

de Jong, Kenneth (1998). Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics, 26,* 283-310.

Golinkoff, Roberta, Kathryn Hirsh-Pasek, Kathleen M. Cauley and Laura Gordon (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23–45.

Herd, Wendy, Allard Jongman and Joan Sereno, (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics, 38,* 504-516.

Horton, Torsten, Frank Bretz and Peter Westfall (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346-363.

Jusczyk, Peter W., and Richard N. Aslin (1995). Infants' detection of the sound pattern of words in fluent speech. *Cognitive Psychology, 23*, 1-23.

Kahn, Daniel (1980). *Syllable-based generalizations in English phonology.* New York: Garland Press.

Mulak, Karen E., Catherine T. Best, Michael D. Tyler and Christine Kitamura (2013). Development of Phonological Constancy: 19-Month-Olds, but Not 15-Month-Olds, Identify Words in a Non-Native Regional Accent. *Child Development*, *84,* 2064-2078.

Pitt, Mark, Laura Dilley and Michael Tat (2011). Exploring the role of variant frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*, 304-311.

Plunkett, Kim (2005). Learning how to be flexible with words. *Attention and Performance, XXI*, 233–248.

Oshika, Beatrice T., Victor W. Zue, Rollin V. Weeks, Helene Neu, and Joseph Aurbach (1975). The role of phonological rules in speech understanding research. *IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-23*, 104-112.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.

Ren, Jie, and James L. Morgan (2011). Sub-Segmental Details In Early Lexical Representation of Consonants. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, August, Hong Kong.

Skoruppa, Katrin, Nivedita Mani, Kim Plunkett, Dominique Cabrol and Sharon Peperkamp (2013). Early word recognition in sentence context: French and English 2-year-olds' sensitivity to sentence-medial mispronunciations and assimilations. *Infancy*, *18*, 1007-1029

Stager, Christine. L. and Janet F. Werker (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381-382.

Swingley, Daniel. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*, 252-269.

Swingley, Daniel, and Richard N. Aslin (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147-166.

Sundara, Megha, Yunjung Kim, James White, and Adam J. Chong (2013). *There is no pat in patting: Acquisition of phonological alternations by English-learning 12-month-olds.* Talk presented at Boston University Conference on Language Development 38, Boston, USA.

Turk, Alice (1992). The American English flapping rule and effect of stress on stop consonant duration. *Working Papers of the Cornell Phonetics Laboratory, 7,* 103-133.

Werker, Janet. F., Christopher T. Fennell, Chris. T., Kathleen M. Corcoran and Christine L. Stager (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy, 3,* 1-30

White, Katherine. S. & James L. Morgan (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114-132.

Zue, Victor, and Martha Laferriere (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America, 66,* 1039-1050.