COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018, Suppl. 2) 606–620

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12528

Developmental Differences Between Children and Adults in the Use of Visual Cues for Segmentation

Ori Lavi-Rotbain,[a] Inbal Arnon[b]

[a]*The Edmond and Lilly Safra Center for Brain Sciences, Hebrew University*
[b]*Department of Psychology, Hebrew University*

Received 31 August 2016; received in revised form 20 May 2017; accepted 22 June 2017

Abstract

Recent work asked if visual cues facilitate word segmentation in adults and infants (Thiessen, 2010). While adults showed better word segmentation when presented with a regular visual cue (consistent mapping between words and objects), infants did not. This difference was attributed to infants' lack of understanding that objects have labels. Alternatively, infants' performance could reflect their difficulty with tracking and integrating multiple multimodal cues. We contrasted these two accounts by looking at the effect of visual cues on word segmentation in adults and across childhood (6–12 years). We found that older children (M_{age} 10;7) benefitted from the regular visual cues, but younger children (M_{age} 7;10), who already knew that objects have labels, did not. Knowing that objects have labels was not enough to use visual cues as an aid for segmentation. These findings show that the ability to integrate multimodal cues develops during childhood, and it is not yet adult-like in children.

Keywords: Statistical learning; Audio–visual input; Word learning; Word segmentation; Language learning; Language development

1. Introduction

Infants, children, and adults are capable of extracting regularities from the environment. This ability—often labeled statistical learning (SL)—is thought to play an important role in language acquisition and learning more generally (e.g., Saffran, Aslin, & Newport, 1996). Most studies of SL tend to focus on learning in one modality (e.g., visual, auditory). However, language presents us with multimodal information: Alongside speech, linguistic interactions contain visual cues like facial expressions, gestures, and the

Correspondence should be sent to Ori Lavi-Rotbain, Psychology Department, The Hebrew University, Mt. Scopus, Jerusalem 9190501, Israel. E-mail: orilavirotbain@gmail.com

objects spoken about. These cues can assist learners in understanding speech and learning linguistic structure. For instance, facial expressions can facilitate adults' segmentation of words in an artificial language (Mitchel & Weiss, 2013). In line with the multimodal nature of language, both infants and adults are capable of integrating multimodal cues during online processing (adults: Cunillera, Càmara, Laine, & Rodríguez-Fornells, 2010a,b; Glicksohn & Cohen, 2013; Mitchel, Christiansen, & Weiss, 2014; Mitchel & Weiss, 2011; infants: Gogate & Bahrick, 1998; Kuhl & Meltzoff, 1982; Teinonen, Aslin, Alku, & Csibra, 2008; Burnham & Dodd, 2004).

The visual presence of an object while its label is heard is one of the more salient visual cues in speech to children (Gogate, Bahrick, & Watson, 2000; Smith, Suanda, & Yu, 2014). Recent work examined the effect of such visual cues on word segmentation in 8-month-old infants and adults to see if the presence of a consistent mapping between objects and labels will facilitate segmentation, and whether it will do so similarly in both infants and adults (Thiessen, 2010). Participants' ability to segment an auditory speech stream (modeled on Saffran et al., 1996) was compared in three conditions. In the *regular visual condition,* a shape appeared on the screen for the duration of each word, serving as a cue to word boundary. Each of the four shapes appeared consistently with one of the four words in the language. In the *irregular visual condition*, the same shapes appeared for the duration of the words, but there was not a consistent mapping: Each shape appeared with all four words. The third condition—the *no-video condition*—included only the auditory stream with no visual cues to segmentation. Infants and adults showed different learning patterns: While adults showed improved segmentation when the mapping was regular, infants did not benefit from this cue and showed the same learning in all three conditions. This pattern was attributed to adults' existing linguistic knowledge that objects have labels: Infants did not show facilitation because they have not yet acquired this knowledge. This was strengthened by two additional findings: (a) Adults showed a positive correlation between segmentation scores and performance on a word-shape correspondence test, and (b) the regular visual cue did not aid segmentation when learning a sequence of tones (where there is not an expectation that tones will be matched to objects). Under this interpretation, young children, who already understand that objects have labels, should perform similarly to adults and show better segmentation in the regular condition compared to the other two.

However, this interpretation is challenged by several recent findings. Importantly, 6-month-old infants show rudimentary lexical understanding, suggesting they already have some knowledge of the association between objects and labels (Bergelson & Swingley, 2012). Additionally, 7-month-olds are already capable of learning arbitrary relations between sounds and objects, an important precursor for word learning (Gogate & Bahrick, 1998). Together, these findings suggest that young infants may already be capable of learning object–label relationships. An alternative explanation to Thiessen's results is that infants' performance reflects their difficulty with tracking and integrating multiple multimodal cues. To benefit from the visual information in the regular condition, learners must be able to track both the statistics of the auditory input and its relation to the visual input, a task that may challenge infants' less developed cognitive resources.

Indeed, recent work shows that limited cognitive resources can lead 11-month-olds to represent probabilistic information differently from adults (Yurovsky, Boyer, Smith, & Yu, 2013). Thus, knowing that objects have labels may not be enough for benefiting from a regular visual cue. If this explanation is correct, we would (a) expect young children to perform similarly to infants and not benefit from visual cues to segmentation and (b) expect to see an improvement with age in the ability to use visual cues to segmentation.

1.1. The current study

In this study, we aim to differentiate between these two accounts by looking at children's use of visual cues for segmentation across childhood (from age 6; 5 to 12 years old). If understanding that objects have labels is enough to benefit from visual cues to segmentation, then children at all the tested ages should show better segmentation in the regular condition. If such knowledge is not enough, the advantage of the regular condition may not be evident in younger children. An additional goal of this study is to further investigate why adults did not benefit from an irregular mapping between shapes and words, even though it provided an additional cue to segmentation (Thiessen, 2010; Glicksohn & Cohen, 2013; but see Cunillera et al., 2010a for a different pattern). It is possible that the benefit of the visual cue to segmentation was over-ridden by the random information that this irregular visual cue introduced. Learners attempt to find a relation between shapes and objects (based on the existence of such a relation in language) where none existed could have interfered with learning the statistics of the language. To see if that is the case, we introduced a new condition—*the heart condition*—where the same shape (a heart) appeared for the duration of each word, then disappeared briefly and re-appeared at the onset of the next word. This condition provides a visual cue to segmentation without random information. If the benefit of the visual cue was over-ridden by the random information in the irregular condition, we should see better segmentation in the heart condition compared to the irregular one. However, if only regular visual cues can facilitate segmentation, we should see no difference between the heart condition and the irregular one.

We address both questions by looking at child and adult use of visual cues for segmentation. In the first study, we tested Hebrew-speaking adults on the three conditions in Thiessen's Experiment 1 and the additional heart condition. This study aims to replicate Thiessen's findings with adults; test the effect of random visual information on performance; and provide an adult baseline for the evaluation of children's performance. The second study examined performance in the same three conditions (without the heart condition) in children between the ages of 6;6–12 to see how age affects the use of visual cues to segmentation.

2. Experiment 1: Adults

In this experiment, we examine adults' segmentation in four conditions: no-video condition, regular condition, irregular condition, and heart condition. We expect performance

in the regular condition to be better than in the no-video condition. If only regular visual cues assist segmentation, then participants will perform better in the regular condition compared to the three other conditions. If, however, temporal visual cues can facilitate segmentation even when they are not regular, then participants will perform better in all three video conditions compared to the no-video condition. In addition, if random information in the irregular condition indeed harmed segmentation, participants should perform better in the heart condition compared to the irregular condition.

2.1. Method

2.1.1. Participants

A total of 138 undergraduate students at the Hebrew University of Jerusalem participated in the study (95 females, 43 males, M_{age} 23;8). All of the participants were native Hebrew speakers without learning disabilities or attention deficits. Participants received 10 NIS or course credit in return for their participation.

2.1.2. Materials

The task was closely modeled on the one used by Thiessen (2010, study 1). All participants were exposed to a familiarization stream containing auditory and visual stimuli. The auditory stream was identical for all participants. There were four different videos, one for each of the different conditions.

2.1.2.1. Auditory stimuli: Auditory stimuli consisted of synthesized speech made up of four unique tri-syllabic words: "*dukame*," "*nalubi*," "*kibeto*," and "*genodi*." The 12 different syllables making up the words were taken from Glicksohn and Cohen (2013). They were created using the PRAAT synthesizer (Boersma & van Heuven, 2001) and were matched on pitch (~76 Hz), volume (~60 dB), and duration (250–350 ms). The four words were created by concatenating the syllables using MATLAB to ensure that there were no co-articulation cues to word boundary. The words were matched for length (average word length 860 ms, range = 845–888 ms). The words were then concatenated together in a semi-randomized order (no word appeared twice in a row) to create the auditory familiarization stream. Transitional probabilities (TPs) between syllables within a word were 1, while the TPs between words were 0.333. The stream was 1:50 min long, and each word repeated 32 times. Importantly, there were no breaks between words and no prosodic or co-articulation cues in the stream to indicate word boundaries.

2.1.2.2. Visual stimuli: In the *no-video* condition, participants saw a static checkerboard image. In the *regular-video* and *irregular-video* conditions, participants saw shapes whose appearance was synchronized with word boundaries. Shapes appeared at word onset and remained onscreen for the duration of the word. In the *regular-video* condition, each word appeared always with the same shape and vice versa ("*dukame*": blue star, "*nalubi*": green hexagon, "*kibeto*": purple heart, and "*genodi*": orange diamond). In *the irregular-video* condition, words and shapes co-occurred randomly with each other, meaning that

Fig. 1. An illustration of the three video conditions.

all shapes appeared with all words by the end of the exposure phase. In the *heart-video* condition, participants saw an image of a purple heart that appeared at word onset and disappeared briefly (for 200 ms) when each word ended: The heart could serve as a visual cue for segmentation. See Fig. 1 for an illustration.

2.1.3. Procedure

Participants completed the experiment on a computer while seated in a quiet room. They were told that after watching the video they would be asked about what they saw and heard. After the exposure phase, participants completed a segmentation test and a word-shape test (only in the regular and irregular-video conditions).

2.1.3.1. Segmentation test: Participants completed 16 two alternative forced choice trials in a random order (with the constraint that the same word/foil did not appear in two consecutive trails). They heard two words and were asked to decide which belonged to the language they heard. The foils ("dunobi," "nabedi," "kilume," and "gekato," average length: 860 ms; range 854–868 ms) were created by taking three syllables from three different words, while keeping their original position in the word. While Thiessen (2010) used part-words as foils, we used non-words. Unlike part-words, non-words never appear together during exposure, making it easier to distinguish between them and real words. Because our ultimate goal was to compare child and adult performance, and because the exposure time in our child study (Experiment 2) was much shorter than previous ASL studies with children (under 3 min compared to over 20 min in Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), we wanted to use more salient distinctions to assess learning (as will be seen, children were far from ceiling even in making the easier distinction between non-words and words). Additionally, since our goal was not to show that adults *can* discriminate words from part-words (a finding shown in other studies), but to see how that ability is impacted by visual cues, we chose to focus only on the "easier" non-word versus word distinction.

Each of the four words appeared once with each of the four foils to create 16 trials. The order of words and foils was counter-balanced so that in half the trials, the real word appeared first and in the other half, the foil appeared first. For subjects in the no-video and heart-video conditions, this was the end of the experiment. Participants in the

regular-video and irregular-video conditions performed an additional word–shape correspondence test.

2.1.3.2. Word–shape correspondence test: This test asked how well participants learned the correspondence between the words and the shapes. In each trial, participants saw the four shapes on the screen and heard one of the four words (see Fig. 2). Then, they had to choose the shape corresponding to the word. Each word was repeated four times on non-consecutive trials, to create 16 trials that appeared in a random order between subjects. This test was performed by participants in the regular-video condition to assess if they had learned the relation between the words and shapes. Participants in the irregular-video condition performed it as well to make sure there were no biases to associate one of the shapes with a certain word (such as in the kiki/bouba effect— Ramachandran & Hubbard, 2001).

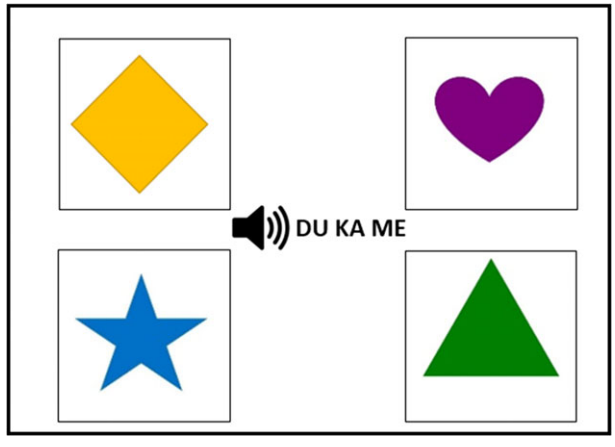

Fig. 2. An illustration of a trial in the correspondence test.

2.2. Results and discussion

Participants were randomly assigned to one of the four experimental conditions. Six participants were excluded from the analyses: four due to technical problems; two due to fatigue. The remaining 132 participants were divided as follows between the four conditions: no-video, 30 participants; regular-video, 36 participants; irregular-video, 33 participants; heart-video, 33 participants. Participants performed best in the *regular-video condition* (M = 79.2%, SD = 14.6%), followed by the *irregular-video* (M = 73.3%, SD = 15.6), *heart-video* (M = 69.7%, SD = 15.7%), and *no-video condition* (M = 64.8%, SD = 14.8%, see Fig. 3). Importantly, participants showed learning in all four conditions: performance differed from chance in all conditions (*regular-video*: $t(35) = 12.02$, $p < .001$; *irregular-video*: $t(32) = 8.6$, $p < .001$; *heart-video*: $t(32) = 7.2$, $p < .001$; *no-video*: $t(29) = 5.5$, $p < .001$).

We used mixed-effect linear regression models to examine the effect of condition on performance (see Jaeger, 2008, for a discussion of the advantage of using mixed-effects

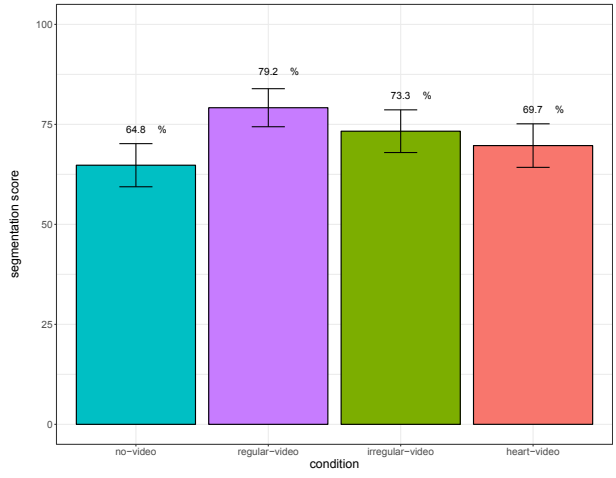

Fig. 3. Adults' mean segmentation score by condition with 95% confidence intervals.

models in analyzing binary responses). Following Barr, Levy, Scheepers, and Tily (2013), the models had the maximal random effect structure justified by the data that would converge. Our dependent binominal variable was accuracy on a single trial of the segmentation test. We had experimental condition (dummy coded, meaning that each condition is compared to the no-video condition) as a fixed effect as well as: trial number (centered); order of appearance in the test (word-first trials vs. foil-first trials) and gender of the participants (male vs. female). The model had random intercepts for participants and items (Table 1). To examine the overall effect of experimental condition we used model comparisons.

As predicted, experimental condition had a significant effect on performance ($\chi(3) = 15.3$, $p < .001$). As found in Thiessen (2010), adults were more accurate in the regular condition compared to the no-video condition ($\beta = 0.79$, $SE = 0.21$, $p < .001$). They were also better in the irregular condition compared to the no-video condition ($\beta = 0.41$, $SE = 0.21$, $p < .05$), a pattern that differs from Thiessen's findings. Performance in the heart condition was no better than the no-video condition ($\beta = 0.22$, $SE = 0.2$, $p > .1$). Trial number significantly affected performance, better accuracy in the beginning of the test ($\beta = -0.03$, $SE = 0.01$, $p < .01$). Order of appearance in the test significantly affected performance, with better accuracy on trials where the word appeared before the foil ($\beta = 0.68$, $SE = 0.1$, $p < .001$). Since the order of presentation of words and foils was counter-balanced this could not reflect a preference for pressing 1 or 2, and is in line with the "interval bias" which is often found in 2AFC tests (Yeshurun, Carrasco, & Maloney, 2008). Gender did not affect performance ($\beta = -0.18$, $SE = 0.11$, $p > .1$). The same pattern was found when we analyzed the results using t tests (the analyses used by Thiessen). Both the regular and irregular conditions were significantly better than the no-video condition ($t(61.48) = 3.95$, $p < .001$; $t(60.89) = 2.22$, $p < .05$, respectively), while the heart condition did not differ from the no-video condition ($t(60.91) = 1.27$, $p > .1$).

Table 1
Mixed-effect regression model for adults in all four conditions

	Estimate	*SE*	*z* value	*p*
(Intercept)	0.51914	0.21886	2.372	**<.05***
Condition (Regular)	0.78953	0.20609	3.831	**<.001*****
Condition (Irregular)	0.41491	0.20679	2.006	**<.05***
Condition (Heart)	0.22153	0.20429	1.084	>.1
Trial number (centered)	−0.03643	0.01116	−3.263	**<.005****
Order of appearance (word)	0.67597	0.10293	6.567	**<.001*****
Gender (female)	−0.22968	0.16119	−1.425	>.1

Note. Variables in bold were significant. Significance obtained using the lmerTest function in R.

2.2.1. Word-shape correspondence results

Participants in the regular condition learned the correct relations between objects and shapes (M = 53.99%, chance = 25%, SD = 32.5%, $t(35) = 5.35$, $p < .001$). However, participants in the irregular condition performed worse than chance, suggesting they had some preference for matchingcertain words with specific objects (M = 20.26%, chance = 25%, SD = 11.7%, $t(32) = -2.33$, $p < .05$). In addition, in the regular condition, we found a positive correlation between the scores in the segmentation test and in the correspondence test: participants who performed better on the segmentation task were also better in selecting the correct label ($r = .43$, $p < .01$). Here, we replicated Thiessen's results with Hebrew-speaking adults showing that the consistent object–label relation facilitated segmentation and word learning.

In sum, we found that Hebrew-speaking adults showed better segmentation when learning an artificial language with regular or irregular visual cues, compared to no visual cues. In addition, participants in the regular-video condition who learned the object–label relation better, showed better segmentation. We did not find a facilitative effect of the heart condition, indicating that a temporal visual cue for word boundaries was not sufficient to improve segmentation, if it was the same cue for all words. These findings also rule out the explanation that the benefit of the visual cue in the irregular condition was over-ridden by the random information found in this condition. Our findings differ from those of Thiessen in one respect: The irregular visual cue also facilitated segmentation compared to no visual cue, a pattern consistent with adults' ability to integrate multimodal cues. The findings on adults' ability to use such a cue is mixed: While some studies do not find facilitation (Glicksohn & Cohen, 2013; Thiessen, 2010), others, using a similar manipulation, have found that an irregular visual cue does benefit learning (Cunillera et al., 2010a). Further work is needed to understand the discrepancy between the studies and the conditions under which an irregular cue can assist segmentation.

3. Experiment 2: Children

Here, we set out to examine our main question: How will visual cues affect children's segmentation? If understanding that objects have labels is enough to be able to use visual

cues to assist segmentation, then children at all the tested ages who already have this linguistic knowledge should perform similarly to adults and show better segmentation in the regular condition. If, however, some processing abilities are necessary as well, then we should see an increase with age in the ability to use the regular visual cue. In particular, young children should not show facilitation in the regular condition compared to the other conditions.

3.1. Method

3.1.1. Participants

A total of 174 children took part in this experiment (age range: from 6;6 to 12 years, M_{age}: 9;3 years; 94 boys, 80 girls). Participants were visitors at the Bloomfield Science Museum in Jerusalem and were recruited for this study as part of their visit to the Living Lab. Parental consent was obtained for all participants. None of the children had known language or learning difficulties and all were native Hebrew speakers. Each child received a small prize for their participation.

3.1.2. Materials

The materials were identical to those used in Experiment 1 minus the heart condition. We compared performance in three conditions: no-video, regular-video, and irregular-video.

3.1.3. Procedure

After receiving parental consent, participants were seated in front of a computer station with a noise-blocking headset next to an experimenter. The children were told they are about to hear an alien language, and that they need to pay attention to what they will see and hear and try to learn it as best as they can. Each child was randomly assigned to one of the three experimental conditions. The procedure for the children was identical to that used in Experiment 1. The instructions were identical in all three conditions.

3.2. Results and discussion

Children were randomly assigned to one of the three experimental conditions. To test the prediction that knowing object–label relations is not enough to benefit from the regular cue, we divided the children into two age groups—"younger" and "older"—based on the median age of the sample (median age: 9;3). Children in both age groups showed learning in all conditions: they were significantly above chance in the three conditions (Table 2). Segmentations scores by experimental conditions and age groups are described in Fig. 4.

We ran two identical mixed-effect models to examine the effect of condition on segmentation in each age group to test the prediction that the younger children will not benefit from any visual cue (like infants) while older children will (more like adults). Our dependent binominal variable was accuracy on a single trial of the segmentation test. We

Table 2
Mean performance and *t* tests of all conditions compared to chance for both age groups

Age Group	Condition	Mean (%)	*SE* (%)	*t* Test	*p*
Younger	No-video (*N* = 32)	54.9	13	$t(31) = 2.13$	<.05*
	Regular-video (*N* = 26)	57.2	12.1	$t(25) = 3.04$	<.01**
	Irregular-video (*N* = 30)	56.9	12.5	$t(29) = 3$	<.01**
Older	No-video (*N* = 29)	57.3	12.7	$t(28) = 3.1$	<.01**
	Regular-video (*N* = 24)	65.1	13.1	$t(23) = 5.63$	<.001***
	Irregular-video (*N* = 33)	60.6	15.3	$t(32) = 3.99$	<.001***

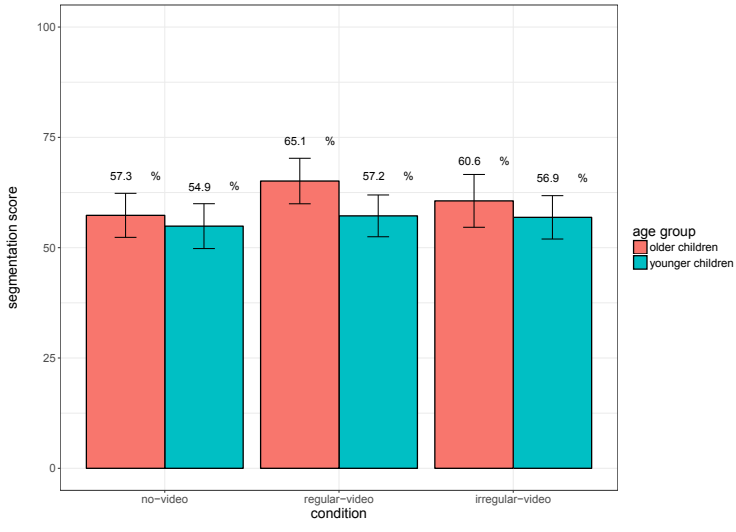

Fig. 4. Children's mean segmentation score by condition and age group.

had experimental condition (with dummy coding, as in Experiment 1) as a fixed effect as well as: age (centered); trial number (centered); order of appearance in the test (word-first trials vs. foil-first trials) and gender of the participants (male vs. female). The model had random intercepts for participants and items (Table 3 for *younger* and Table 4 for *older*). To examine the overall effect of experimental condition, we used model comparisons.

As predicted, the younger group looked like the infants in Thiessen's study: They did not perform better in the regular condition despite knowing that objects have labels ($\beta = 0.08$, $SE = 0.14$, $p > .5$; $\chi(2) = 0.09$, $p > .5$). Performance in the irregular-video condition did not differ from the no-video condition ($\beta = 0.06$, $SE = 0.13$, $p > .5$). Younger children showed better accuracy on trials where the word appeared before the foil ($\beta = 0.45$, $SE = 0.11$, $p < .001$), but unlike adults they were not affected by trial number ($\beta = -0.01$, $SE = 0.01$, $p > .1$). The same pattern was found using *t* test: performance did not differ between the no-video condition and the two visual conditions (regular vs. no-video: $t(54.91) = 0.71$, $p > .1$; irregular vs. no-video: $t(59.94) = 0.61$, $p > .52$).

Table 3
Mixed-effect regression model for younger children in all four conditions

	Estimate	*SE*	*z* value	*p*
(Intercept)	0.27369	0.14937	1.832	0.07
Age (centered)	0.15208	0.06214	2.447	**<.05***
Condition (Regular)	0.08807	0.13523	0.651	>.5
Condition (Irregular)	0.06640	0.12923	0.514	>.5
Trial number (centered)	−0.01005	0.01177	−0.854	>.1
Order of appearance (word)	0.44571	0.10854	4.106	**<.001*****
Gender (female)	−0.16025	0.10937	−1.465	>.1

Note. Significance obtained using the lmerTest function in R.

Table 4
Mixed-effect regression model for older children in all four conditions

	Estimate	*SE*	*z* value	*p*
(Intercept)	−0.144527	0.180186	−0.802	>.1
Age (centered)	0.136201	0.082388	1.653	0.098
Condition (Regular)	0.348871	0.163649	2.132	**<.05***
Condition (Irregular)	0.182354	0.151060	1.207	>0.1
Trial number (centered)	−0.044989	0.012387	−3.632	**<.001*****
Order of appearance (word)	0.490969	0.12323	4.327	**<.001*****
Gender (female)	−0.001422	0.137145	0.010	>.5

Note. Significance obtained using the lmerTest function in R.

In contrast, the older group performed more similarly to adults and showed better segmentation when a regular visual cue was present: They were better in the regular condition compared to the no-video one ($\beta = 0.35$, $SE = 0.16$, $p < .05$). Unlike our adult sample, they were not aided by the irregular visual cue ($\beta = 0.18$, $SE = 0.15$, $p > .1$). Like adults, older children showed better accuracy on trials where the word appeared before the foil ($\beta = 0.49$, $SE = 0.12$, $p < .001$), but they were negatively affected by trial number ($\beta = -0.04$, $SE = .01$, $p < .001$). The t tests revealed the same pattern (regular vs. no-video: $t(48.52) = 2.17$, $p < .05$; irregular vs. no-video: $t(59.85) = 0.92$, $p > .1$).

3.2.1. Word-shape correspondence results

In line with the adult findings, older children did manage to learn the object–label mappings ($M = 34.4\%$, chance = 25%, $SD = 21.6\%$, $t(23) = 2.12$, $p < .05$). Younger children, like the infants in Thiessen's study, did not manage to learn the mappings ($M = 25.96\%$, $SD = 12.9\%$, $t(25) = 0.3$, $p > .5$) even though they know that objects have labels. As expected, children in the irregular condition in both age groups performed at chance (younger: $M = 21.88\%$, $SD = 12.8\%$, $t(29) = -1.34$, $p > .1$; older: $M = 23.86\%$, $SD = 14.9\%$, $t(32) = -0.44$, $p > .5$).

In sum, in Experiment 2, we found that only older children showed better segmentation in the regular condition compared to the no-video condition. Since younger children already have the understating that objects have labels, and still did not show facilitation, the findings are more consistent with the explanation that object–label relations are not enough to benefit from visual cues. The older children still differed from adults in that they did not benefit from an irregular visual cue. This pattern may reflect their more limited ability to use inconsistent information. However, since the adult findings on the use of such a cue are also mixed, it is too early to draw developmental conclusions from this pattern.

4. Discussion

We set out to distinguish between two explanations for the facilitative effect of regular visual cues on segmentation: Is the effect driven by the knowledge that objects have labels or does it reflect the difficulty with integrating multiple multimodal cues? To do so, we examined word segmentation in an artificial language with and without visual cues in both children and adults. As in Thiessen (2010), our adult sample showed better segmentation when regular visual cues were present. We expanded on those findings by ruling out the explanation that random information in the irregular condition interfered with learning. Participants showed enhanced segmentation in the irregular condition, but not in the heart condition (where there was less random information since the same visual cue appeared with all words). Interestingly, unlike Thiessen, we did find that segmentation improved in the presence of an irregular visual cue. This pattern mirrors that found by Cunillera et al. (2010a), who used a very similar manipulation and found that having an irregular visual cue facilitated segmentation in adults. Interestingly, a similar trend is present in Thiessen's results (regular cue: 75%, irregular cue: 61%, no-video: 55.5%). The lack of a significant effect in Thiessen's study may have also been driven by the use of less sensitive *t* tests to analyze the results. That is, having an irregular cue may indeed facilitate segmentation. Such a pattern fits in with the processing perspective we offer: In the irregular condition there is a visual cue to segmentation even if the mapping between the visual cue and the words is not consistent.

Developmentally, we found that younger children (M_{age} = 7;10) performed like infants: Despite knowing that objects have labels, they did not benefit from regular visual cues for segmentation. In contrast, older children (M_{age} = 10;7) performed similar to adults: They showed better segmentation in the regular condition. Unlike adults, they did not benefit from the irregular condition. Together, these findings suggest that the ability to use regular visual cues is not driven only by the linguistic knowledge that objects have labels, a pattern consistent with findings showing that infants already have some rudimentary object–label associations (Bergelson & Swingley, 2012; Gogate & Bahrick, 1998). Instead, they suggest that the ability to integrate multimodal cues develops during childhood and is not yet adult-like in both infants and young children. This pattern is in line with recent work suggesting age-related changes in SL during childhood (Arciuli &

Simpson, 2011; Raviv & Arnon, in press; Saffran et al., 1997). It is possible that younger children may be able to use visual cues if the task were made easier. However, such a finding would further strengthen the interpretation that the inability to use visual cues for segmentation is related to processing difficulty and not to object–label knowledge. Moreover, our task was already easier than that used in Thiessen since we used non-words as foils (and not part-words). Even under these easier conditions, younger children did not show better performance in the presence of visual cues.

How, then, are we to reconcile these findings with the large literature showing that infants are capable of integrating multiple cues? Interestingly, much of this work focuses either on the integration of multiple auditory cues or on the integration of linguistic visual cues (e.g., lip movement, facial expression) with auditory ones. Many studies show that infants are capable of simultaneously attending to different auditory cues. For instance, in studies of word segmentation, young infants attend more to speech cues (stress and co-articulation) than to the TPs between syllables (another auditory cue), though they can compute both (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). Moreover, they are also capable of integrating prosodic and phonotactic cues to word segmentation (Mattys, Jusczyk, Luce, & Morgan, 1999). Similarly, infants are capable of integrating bimodal cues to speech as evidenced by the fact that they show the McGurk effect (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997), and that the magnitude of the effect is increased when they are also exposed to a visual articulation of the syllables (Teinonen et al., 2008).

However, there is less work that examined infants' ability to integrate non-linguistic visual cues (such as a presence of an object) with auditory linguistic cues when those are not presented as segmented words. The work on cross-situational learning shows that infants can compute cross-situational statistics of both objects and labels (e.g., Smith & Yu, 2008). Similarly, Gogate and colleagues show that even younger infants can map sounds to objects if those appear in synchrony (Gogate & Bahrick, 1998, 2001). In both cases, labels are already segmented words. In contrast, both Thiessen's results and the ones reported here indicate that the ability to use objects as visual cues for segmentation takes time to develop. Taken together, the findings suggest that not all cues are equally available: While the basic ability to integrate multimodal cues is present from infancy, there are developmental changes in the kind of cues that can be successfully integrated.

Acknowledgments

We wish to thank Noam Siegelman for his help with the statistical analyses, as well as the research assistants at the Living Lab, the parents and children who participated, and the Bloomfield Science Museum. The research was funded by the Israeli Science Foundation grant number 584/16 awarded to the second author.

References

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109.

Boersma, P., & van Heuven, V. (2001). Speak and unspeak with praat. *Glot International*, *5*(9–10), 341–347.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. https://doi.org/10.1002/dev.20032.

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010a). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology (2006)*, *63*(2), 260–274. https://doi.org/10.1080/17470210902888809.

Cunillera, T., Laine, M., Càmara, E., & Rodríguez-Fornells, A. (2010b). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, *63*(3), 295–305.

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*(6), 1161–1169. https://doi.org/10.3758/s13423-013-0458-4.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438.

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy*, *2*, 219–231. https://doi.org/10.1207/S15327078IN0202_7.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567. https://doi.org/10.1006/jmla.2000.2755.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. https://doi.org/10.1006/cogp.1999.0721.

Mitchel, A. D., Christiansen, M. H., & Weiss, D. J. (2014). Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Frontiers in Psychology*, *5*(May), 1–6. https://doi.org/10.3389/fpsyg.2014.00407.

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1081–1091. https://doi.org/10.1037/a0023700.

Mitchel, A. D., & Weiss, D. J. (2013). Visual speech segmentation: Using facial cues to locate word boundaries in continuous speech. *Language and Cognitive Processes*, *May*, 1–10. https://doi.org/10.1080/01690965.2013.791703.

Ramachandran, S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Raviv, L., & Arnon, I. (in press). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developemental Science*.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357. https://doi.org/10.3758/BF03211902.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. a., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the comer of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*(6), 1093–1106. https://doi.org/10.1111/j.1551-6709.2010.01118.x.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706.

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *48*(17), 1837–1851. https://doi.org/10.1016/j.visres.2008.05.008.

Yurovsky, D., Boyer, T. W., Smith, L. B., & Yu, C. (2013). Probabilistic cue combination: Less is more. *Developmental Science*, *16*(2), 149–158. https://doi.org/10.1111/desc.12011.