Perceptual learning of intonation: Adults are more narrow-minded than children
Vsevolod Kapatsinski, Paul Olejarczuk, Melissa A. Redford
University of Oregon

ABSTRACT: Perceptual learning refers to the process of mapping points in the continuous acoustic similarity space onto linguistic categories. The present paper reports on rapid perceptual learning of intonation contour categories in children and adults. We document age-related differences in inductive bias. Given the same experience, adults learn narrower categories than children, which may make it harder for them to recognize new instances of a contour as belonging to the same category. A major transition is reported around age 10: 11-year-olds are largely adult-like but 9-year-olds are not. This finding may help explain why children are better second language learners than adults.

Keywords: second language acquisition, sensitive period, prosody, intonation, perceptual learning, inductive bias

1. Introduction

One of the central challenges for language acquisition is to explain why children are better than adults at learning languages. Two contradictory explanations have been advanced. One, the "less is more" (Newport 1990) or "starting small" (Elman 1993) hypothesis, suggests that children have helpful cognitive limitations, including a smaller working memory capacity. These limitations help children focus on narrower temporal spans of speech, allowing them to analyze the incoming signal in greater detail, albeit at the expense of temporal integration. The other explanation for children's superior language learning, the “starting big” hypothesis (Arnon 2009, Arnon & Ramscar 2012; see also Tomasello 2003), suggests the exact opposite: "children learn from units that are larger and less analyzed than the ones adults learn from". Both hypotheses have been applied to the acquisition of morphosyntax. A principal aim of the present study is to evaluate their applicability to the acquisition of prosody, specifically, intonation contours.

Like some syntactic constructions, intonation contours are extended in time. This characteristic makes them prime candidates for testing between the two hypotheses in a new domain, namely perceptual learning of intonation contour categories. So far, work in perceptual learning has focused on the acquisition of phoneme categories in infancy (Maye et al. 2002) and on the fine-tuning of these categories in adults (Bertelson et al. 2003, Idemaru & Holt 2011, Norris et al. 2003). Thus the present study extends work on perceptual learning from the segmental to the suprasegmental level.

We define perceptual learning as learning to categorize locations in a continuous multidimensional perceptual similarity space. Part of this process is acquiring which features are especially important to a category (e.g. Idemaru & Holt 2011, Maye et al. 2008). The features that define intonation contours are the relative pitch values (Low or High) of the contour’s inflection points (Hayes 2009:295-302, Pierrehumbert & Hirschberg 1990), which are referred to as tones (L or H). The slopes between the points may also be important (Cho 2010, ’t Hart et al. 1990). Accordingly, the category is either defined as a sequence of tones (inflection points) and/or as a sequence of pitch movements (slopes). Evidence for categories of entire contours comes from the idiomatic value of some contours such as the so-called surprise-redundancy contour (Liberman 1978), which is described in modern theory as the tonal sequence L* H* L-L% (Beckman & Ayers 1997), where * represents tones aligned to stressed syllables in critical words and % represents tones aligned with a phrase boundary.

Intonation contour categories are language-specific, and therefore must be learned. For example, Grabe et al. (2003) show that some intonation contours are categorized differently by English,
Spanish and Mandarin speakers. Furthermore, the cross-linguistic differences in contour categorization disappear if the contours are superimposed on non-linguistic stimuli, suggesting that the categories are based on linguistic experience.

As discussed in Mitchell (1980), unbiased learning of categories from positive examples is impossible. Every set of examples of category members is consistent with a range of hypotheses varying between the maximally general hypothesis "everything is in the category" and the maximally specific hypothesis "only the experienced members are in the category". Different rational learners may therefore assume categories of different levels of generality. If children "start small", their categories are expected to be broader than those of adults: they may focus on only some of the features characterizing the contour, rather than considering the entire sequence of features. In addition, if their memory representations of the experienced contours are underspecified compared to those of adults, they may be more tolerant of mismatch between new examples and stored members of the category. If, instead, they “start big,” storing particular exemplars of intonation contours and not generalizing as much as adults do, their categories are expected to be narrower than those of adults.

Children and adults were exposed to different intonation contours. We tested for differences in category breadth by testing participants on novel exemplars that were generated from the same prototype as the training exemplars, but deviated from the prototype as much as or more than the training exemplars. In addition, we examined whether children, like adults, attend to inflection points and slopes rather than absolute pitch level, by creating a flat contour prototype. Our reasoning was that if slopes or inflection points are relevant to the categorization of contours, then participants should treat the flat prototype differently than exemplars with any kind of pitch movement, even if the exemplar pitch contours average out to the flat prototype. On a nativist view of language acquisition (Chomsky 1981), the bias to attend to inflection points / slopes, being true of all languages, is a prime candidate for an innate principle. Nativist explanations of children's superior language learning abilities rely on assuming that children have access to knowledge of such innate principles and that access to this knowledge gradually weakens with development (e.g., Johnson & Newport 1989, Lenneberg 1967). If this account is true, we would predict the bias to define contours using slopes or inflection points to weaken as children become older. On the other hand, the bias might be strengthened by experiencing that intonation contours in the L1 are defined by inflection points and slopes (as argued for another putative universal, the old-before-new word order, by Narasimhan & Dimroth 2008). In that case, the bias is expected to strengthen with age.

2. Methods and Materials

2.1. Experiment Design

The experiment consisted of a training stage followed by a test stage. The training stage was split into multiple blocks. Each of the blocks involved presenting participants auditorily with small deviations from a prototype intonation contour. Each example of a contour was paired with a meaning, a labeled picture of an alien who "talks like this." Following Xu & Tenenbaum (2007), examples were blocked by category in order to avoid having the participants focus exclusively on the features that distinguished the presented categories. Block order was randomized. In the test stage, participants heard new and familiar examples of the experienced categories and saw the referents of all trained categories on the screen at the same time. They also saw a picture of many other aliens. They were asked to decide whether what they were hearing was said by one of the familiar pictured aliens (and if so, which one) or by some alien they had not seen or heard before ("Is this dax, wug, swinkle, or one of these other guys?"). They responded by clicking on the appropriate picture. The additional 'garbage bin' response option ("one of these other guys") allowed us to distinguish between confusability, p(\text{Response}=\text{category}_i \mid \text{Response} \neq \text{garbage})

and acceptability, p(\text{Response}=\text{category}_i \mid \text{Response}=\text{garbage}").
Response=(category, OR 'garbage') of the various category exemplars presented in the test stage. Since we were primarily interested in category breadth, it is acceptability that we analyze in the present report.

2.2. Stimuli

Intonation contours were synthesized in Praat (Boersma 2001), and consisted of random deviations around three prototypes: a 'flat' contour: -----, a 'final-fall' contour: \_\_\_, and an 'M' contour: \_/\_. The contours were superimposed onto identical 16-syllable [mimimi] sequences, created from a recorded utterance of a male speaker of U.S. English (Weinberger 2013) by resynthesis and phoneme replacement in MBROLA (Dutoit et al. 1996). All inflection points were aligned with vowel onsets except the second peak of the 'M', which was placed at vowel offset. To generate training exemplars, the pitch level of every inflection point and every vowel midpoint in the prototype was randomly perturbed up or down by one semitone or left unchanged. For each exemplar, an inverse copy with the opposite perturbation pattern was created, so that the pair averaged out to the prototype. All together, six such pairs of low-level distortions were made for a total of twelve training exemplars per category. The test stimuli included 1) four of the trained exemplars, 2) four new low-level distortions created in the same manner, 3) the prototypes, which the participants were not trained on, 4) four each intermediate- and high-level distortions, wherein each syllable deviated from the prototype by up to 3 or 5 semitones, respectively, and 5) for the multi-featured contour \_/\_, distractor examples that lacked one of the features, specifically, four with only an early peak \____, four with only a late peak ___\_, and four 'hat' contours that lacked the medial fall-rise sequence, /\_/\_. The pitch contours are shown in Figure 1.

Figure 1. Pitch contours used in testing. Black dashed line = prototypes. Magenta = one-semitone distortions (four of which were used in training) and, for 'M', the distractors. Blue = three-semitone distortions. Green = five-semitone distortions.

2.3. Participants

40 children and 42 adults participated in the experiment. Adults were all native American-English speakers with normal hearing recruited from introductory psychology classes through the Psychology/Linguistics Human Subject Pool. Children were native American-English speakers, who were
recruited from the local community. Typical development was confirmed by parental report; normal hearing with a pure tone hearing screen. 16 children were 9 years old at the time of study, 15 were 10 years old, and 9 were 11 years old. The median age was 10;3.

2.4. Statistical analysis
Statistical analyses were performed using mixed-effect logistic regression as implemented in the lme4 package (Bates & Maechler 2010) in R (R Development Core Team 2010). We used participants and items as crossed random effects and included within-participant random slopes for within-participant predictors: category and distortion level. We included within-item random slopes for age. Age was categorized into three categories: younger children (younger than 10;3), older children, and adults. Error bars in the graphs are 95% confidence intervals based on a proportion test (prop.test() in R, Newcombe 1998, Wilson 1927).

3. Results
Figure 2 shows the extent to which \[\_\_\], \_\_\_ and \[\_\_\_\_\] were classified as instances of \[\_\_\] rather than as garbage. Younger children accepted \_\_\_ and \_\_\_ as instances of \[\_\_\] whereas older children and adults did not (the age category by stimulus category interaction was significant: younger children accepted all three distractor types as much as actual examples of \[\_\_\_\] ; older children did not differ from adults (see Table 1 for the statistics). Furthermore, novel low-distortion exemplars of \[\_\_\_\_] were categorized as \[\_\_\_\_] more often than the distractors by adults and older children, but not by younger children (see the confidence intervals in Figure 2). These findings suggest that younger children have formed broader categories than older children and adults, as predicted by the "starting small" hypothesis.

Figure 2. Acceptance of early peak (\[\_\_\]), late peak (\_\_\_), and hat (\[\_\_\_\_\]) distractors and new low-level (1 semitone) distortions of 'M' (\[\_\_\_\]) by age group. Error bars are 95% confidence intervals.
Table 1. Fixed effect estimates and significance levels for the data in Figure 2. 'adults' and 'ᶠá́' are reference levels.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>older children</td>
<td>-0.7441</td>
<td>-0.748</td>
<td>0.4544</td>
</tr>
<tr>
<td>younger children</td>
<td>-2.8376</td>
<td>-2.853</td>
<td>0.0043</td>
</tr>
<tr>
<td>⁾</td>
<td>-4.6614</td>
<td>-7.743</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>/~~~</td>
<td>-5.633</td>
<td>-8.575</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>older children : ⁾</td>
<td>0.3089</td>
<td>0.319</td>
<td>0.7501</td>
</tr>
<tr>
<td>younger children : ⁾</td>
<td>4.5626</td>
<td>4.699</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>older children : /~~~</td>
<td>0.1176</td>
<td>0.136</td>
<td>0.8914</td>
</tr>
<tr>
<td>younger children : /~~~</td>
<td>2.8189</td>
<td>3.271</td>
<td>0.0011</td>
</tr>
<tr>
<td>older children : /~~~</td>
<td>-0.102</td>
<td>-0.093</td>
<td>0.9260</td>
</tr>
<tr>
<td>younger children : /~~~</td>
<td>3.5841</td>
<td>3.342</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

Figure 3 shows that all distortions (of any category) that were further away from the category prototype than previously experienced exemplars were more likely to be accepted by younger children than by older children and adults (see Table 2 for the statistics). This was true for all contours. The older children were again adultlike in their performance. These results are again consistent with the "starting small" hypothesis, and with a change in inductive bias around 10;3.

Figure 3. Acceptance of deviations by distortion level and age group. (TR1 = previously experienced exemplars, ST = semitones by which the exemplars can maximally deviate from the prototype). Error bars are 95% confidence intervals.
Table 2. Fixed effect estimates and significance levels for the data in Figure 3. 'adults' and '1st' are reference levels.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 st</td>
<td>-2.15023</td>
<td>-7.689</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>5 st</td>
<td>-3.8528</td>
<td>-11.905</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>older children</td>
<td>-0.05793</td>
<td>-0.157</td>
<td>0.8749</td>
</tr>
<tr>
<td>younger children</td>
<td>-0.78126</td>
<td>-1.899</td>
<td>0.0575</td>
</tr>
<tr>
<td>3 st : older children</td>
<td>0.09333</td>
<td>0.266</td>
<td>0.7905</td>
</tr>
<tr>
<td>5 st : older children</td>
<td>0.61575</td>
<td>1.412</td>
<td>0.1580</td>
</tr>
<tr>
<td>3 st: younger children</td>
<td>1.84557</td>
<td>4.495</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>5 st: younger children</td>
<td>2.0721</td>
<td>4.198</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Finally, Figure 4 shows that all participant groups tended to reject the flat prototype as a member of the category formed from it. This was not the case for the prototypes of the other contour categories, which differed from the flat prototype in having characteristic slopes and inflection points that were preserved in the exemplars formed from them (Figure 5). This result is consistent with the consensus opinion that inflection points and/or slopes are the important features of intonation contours. However, children were more likely to accept the flat prototype than adults. In this case, the younger and older children patterned together, suggesting that the developmental timecourse for rejecting the flat prototype was different than the timecourse for rejecting massive (3-semitone and 5-semitone) distortions of the contour (Figure 3) or distractors that lacked defining features of the experienced category (Figure 2).

Figure 4. Acceptance of prototypes (P), previously experienced exemplars (TR1), and new low-level distortions (1ST) for the flat category. Error bars are 95% confidence intervals.
Figure 5. Acceptance of prototypes (P), previously experienced exemplars (TR1), and new low-level distortions (1ST) for the 'M' and final-fall categories. Error bars are 95% confidence intervals.

Table 3. Fixed effect estimates and significance levels for the data in Figure 4. 'Adults' and 'TR1' are reference levels.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prototype</td>
<td>-3.4018</td>
<td>-7.792</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>1ST</td>
<td>-0.1484</td>
<td>-0.472</td>
<td>0.6372</td>
</tr>
<tr>
<td>older children</td>
<td>-0.4273</td>
<td>-0.832</td>
<td>0.4055</td>
</tr>
<tr>
<td>younger children</td>
<td>-0.8916</td>
<td>-1.47</td>
<td>0.1415</td>
</tr>
<tr>
<td>Prototype: older children</td>
<td>1.6268</td>
<td>2.395</td>
<td>0.0166</td>
</tr>
<tr>
<td>1ST: older children</td>
<td>0.1413</td>
<td>0.287</td>
<td>0.7743</td>
</tr>
<tr>
<td>Prototype: younger children</td>
<td>1.7408</td>
<td>2.179</td>
<td>0.0293</td>
</tr>
<tr>
<td>1ST: younger children</td>
<td>0.4333</td>
<td>0.702</td>
<td>0.4826</td>
</tr>
</tbody>
</table>

4. Discussion

Younger children form broader categories than older children and adults when exposed to novel intonation contours. This finding is consistent with the "starting small" hypothesis (Elman 1993, Newport 1990), and not with the "starting big" hypothesis (Arnon 2009, Arnon & Ramscar 2012). Additionally, it appears that neither children nor adults rely on their memory of specific exemplars to determine category membership. In fact, no significant advantage was observed for familiar exemplars compared to new exemplars for any age group or category; cf. 1ST (new) vs. TR1 (familiar) in Figures 4 and 5. This result contrasts with Posner & Keele's (1968) classic findings for simple visual patterns and suggests relatively abstract representation of intonation patterns.

The finding of greater breadth of children's categories is in line with work on the development of concepts (Clark 1973, Mandler 2000, Pauen 2002, Rogers & McClelland 2004). Children have also been found to be tolerant of mispronunciations of unfamiliar words (Fennell & Werker 2003, Swingley
Why are children so broad-minded? One explanation is that their memories of prior experiences is more "blurry", and thus more underspecified, than adult memories (Newport 1990). Another explanation is that children have not yet formed as many categories in their native language as adults, and hence may underestimate the number of distinct categories "out there." Whereas an adult might be tempted to quip that "There are more things in Heaven and Earth, Horatio, than are dreamt of in your philosophy" (Shakespeare 1603), less may truly be more when it comes to language acquisition (Elman 1993, Newport 1990). Rogers & McClelland (2004:96) argue that "there may be good computational reasons ... to begin with ... undifferentiated internal representations. Specifically, such an initial state would permit very rapid learning about properties common to all things", or in our case, all intonation contours. Broader categories may also be helpful for dealing with variability in production and perception (Logan et al. 1991). And it may be that children have to manage more variability than adults, particularly in the process of matching their own productions to the adult target. Immature motor control certainly results in temporal variability during the articulation of speech segments and syllables (Kent & Forner 1980, Lee et al. 1999), and so may contribute to variability in pitch patterns during the production of intonation contours.

Another important finding of the present study is that the unnatural flat prototype was rejected as being a member of its category by both children and adults. This is presumably because the prototype did not share inflection points or pitch slopes with the exemplars created from it. For adults, the flat prototype was as bad as a 5-semitone distortion. The bias against the unnatural prototype appears to grow with development, presumably due to increasing language experience. Despite attention to relative heights of inflection points and slopes being universal (see Narasimhan & Dimroth 2008 for another example), the bias does not weaken with maturation contra Chomsky (1981) and Newport & Johnson (1989).

A final, unexpected finding was that the late peak was more important than the early peak for category membership. The late peak may simply have been more perceptually salient than the early peak because participants may have compensated for declination, which describes the natural decline in pitch across an utterance (Pierrehumbert 1979). The declination in our stimuli may not have been steep enough for listeners to perceptually equalize peak heights. A more interesting possibility is that adults learn to preferentially attend to the end of the utterance because it usually is the most informative part (Clark & Clark 1978, Wundt 1900). This possibility fits with Narasimhan & Dimroth's (2008) finding that children, having not yet learned that new information comes at the end of the sentence, tend to place new information at the beginning. If this interpretation of the result is correct, then the finding is yet another example of a linguistic universal that is the outcome of, rather than prerequisite to, language learning.

5. Conclusion
Children and adults rapidly learn new intonation categories. However, the categories they form differ. Given the same experience, 9-year-olds form broader categories than adults while 11-year-olds’ categories are adultlike in breadth. However, even 11-year-olds differ from adults when it comes to being focused on the inflection points or slope values of a contour. Despite being universal, attention to the relative pitch of inflection points strengthens with language experience.

Acknowledgements
This research was supported by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily reflect the views of NICHD.
References


