Perceptual learning of intonation contour categories in adults and 9 to 11-year-old children: Adults are more narrow-minded

Vsevolod Kapatsinski, Paul Olejarczuk & Melissa A. Redford.

University of Oregon

ABSTRACT: The present paper reports on rapid perceptual learning of intonation contour categories in 9- to 11-year-old children and adults. Intonation contours are temporally-extended patterns whose perception requires temporal integration and is therefore expected to pose significant working memory challenges. We report that both children and adults form relatively abstract representations of intonation contours: previously encountered and novel exemplars are categorized together equally often, as long as distance from the prototype is controlled. However, age-related differences in categorization performance also exist. Given the same experience, adults form narrower categories than children. In addition, adults pay more attention to the end of the contour while children appear to pay equal attention to the beginning and the end. The age range we examine appears to capture the tail-end of the developmental trajectory for learning intonation contour categories: there is a continuous effect of age on category breadth within the child group, but the oldest children (older than 10;3) are largely adult-like.

Keywords: prosody, intonation, categorization, perceptual learning

1. Introduction

The present paper explores perceptual learning of intonation contour categories in children and adults. Perceptual learning is a type of category learning where the learned categories are thought to be useful for future perceptual processing (e.g. Goldstone, 1998). So far, work on the perceptual learning of language has focused mainly on the acquisition of phoneme categories in infancy (Maye et al., 2002) and on the fine-tuning of these categories in adults (e.g. Bertelson et al., 2003; Idemaru & Holt, 2011; McQueen et al., 2006; Norris et al., 2003; Nygaard & Pisoni, 1998; Pisoni et al., 1994; Reinisch et al., 2014). Recent work has investigated perceptual learning of local suprasegmental patterns: lexical tone distinctions instantiated over single syllables (Francis et al., 2008; Maddox et al., 2013; Mitterer et al., 2011), as well as pitch accents and stress patterns instantiated over pairs of adjacent syllables (Shport, 2011; Reinisch & Weber, 2012). Work on non-linguistic perceptual category learning has likewise focused on stimuli with little to no temporal extent (static images, as in Gibson & Gibson, 1955; Posner & Keele, 1968; Shepard et al., 1961, *et seq*). We are in fact aware of only two studies that have studied how development affects acquisition of categories comprising temporally extended patterns (Berger & Hatwell, 1996, on haptically explored objects, and Schwarzer, 1997, on melodies). The present study thus extends work on perceptual learning by investigating the acquisition of novel temporally-extended suprasegmental patterns, namely meaningful intonation contour categories (see also Kurumada et al., 2012). In this introductory section, we aim to convince the reader that perceptual learning of intonation contours, and temporally extended patterns more generally, is an interesting research problem for students of categorization, especially those interested in how category learning biases change over the course of development.

1.1.Intonation contour categories

Intonation contours are meaningful pitch patterns distributed over utterances (e.g. ‘t Hart et al., 1990). For example, changing the intonation contour of an utterance can change a statement into a question. Intonation contours can also provide pragmatic information. Thus the surprise redundancy contour shown in Fig. 1 conveys a kind of incredulous disapproval (Sag & Liberman, 1975). Idiomatic contours like the surprise-redundancy contour suggest that language learners can acquire meaningful categories of entire contours.

Intonation contours are usually described by the relative pitch values (Low or High) of the contour's inflection points (tonal targets, Bruce & Gårding, 1978; Ladd, 1983; Liberman, 1978; Pierrehumbert, 1980, 2000; Pike, 1945). Accordingly, a contour category can be defined as a sequence of features (=tonal targets) that are shared in most instances of the contour, and occur in a necessary order (see also ‘t Hart et al., 1990, pp.82-88). For example, the surprise-redundancy contour shown in Fig. 1 is described as the sequence H L* H* L% (e.g. Hayes, 1995, pp.16-18), where H refers to a high tone, L to a low tone, * represents alignment to the stressed syllables in critical words, and % indicates alignment with a phrase boundary. As one can see in Fig. 1, the temporal intervals between tonal features are variable and can be quite long.

FIGURE 1 ABOUT HERE

There is disagreement in the literature as to whether intonation contours are monomorphemic (e.g. Sag & Liberman, 1975), or whether they are sequences of simpler tonal morphemes (e.g. Gussenhoven, 2004; Pierrehumbert & Hirschberg, 1990). The compositional view of intonation may lead some to wonder whether intonation contours are plausibly thought of as perceptual categories. However, research on morphological processing suggests that words and even phrases composed of multiple morphemes are often perceived as single units (e.g. Baayen et al., 1997; Sereno & Jongman, 1997; Tremblay, 2009). For example, Tremblay (2009) showed that frequency of a compositional four-word phrase like *in the middle of* influences processing as early as lexical influences can be detected in the ERP signal, suggesting that predictive processing in comprehension proceeds at multiple levels in parallel. In other words, the listener appears to hypothesize *in the middle of* as early as the beginning of *in* and would therefore profit from learning the acoustic characteristics of the category of manifestations of *in the middle of*. Based on this work, we suggest that even if an intonation contour is not a morpheme – i.e., if it contains smaller meaningful parts – instances of a contour are still likely to form a perceptual category (see also Gussenhoven, 2004; ‘t Hart et al., 1990).

1.2. Intonation contour perception and working memory

Intonation contours exhibit structure that has been underexplored in the literature on category learning. For this reason, studying the perceptual learning of contour categories in children and adults has the potential to shed light on important questions in the field. In particular, intonation contours offer us the opportunity to study the relationship between categorization and working memory. In order to perceive an intonation contour, one must retain each tonal feature as it is presented in the speech stream and then integrate the accumulated features into a larger structure. This process of *temporal integration* has been proposed to be a major function of working memory (Halford et al., 1998; Oberauer et al., 2000, 2003; Robin & Holyoak, 1995). Prior studies that have looked at the relationship between categorization and working memory have used stimuli that require little or no temporal integration (Lewandowsky et al., 2012; Minda et al., 2008; Rabi & Minda, 2014a). The present study on perceptual learning of intonation contour categories thus contributes to understanding the role of working memory in category formation by using temporally extended stimuli.

As noted by many researchers, category learning biases may change over the course of development (e.g. Aslin & L. Smith, 1988; L. Smith, 1989; Ward et al., 1990). Developmental changes in category learning are especially likely for categories of temporally extended stimuli because working memory capacity is known to continue developing into adolescence (e.g. Luna et al. 2004). An individual with low working memory capacity may have trouble keeping track of and integrating all necessary features of the category, resulting in lower-dimensional categories. Since a category with fewer necessary features is less specific than one with more necessary features, a low-dimensional category will be broader than a high-dimensional category.

Indeed, developmental studies indicate that children's categories are often low dimensional compared to adult categories. When stimuli are characterized by multiple individually informative and mutually redundant features (each feature affording a near-perfect category discrimination), children tend to focus on only one feature, whereas adults are more likely to attend to several features at once (Thompson, 1994; Ward & Scott, 1987; Ward et al., 1990). Children's low dimensional categories are likely explained by their smaller working memory capacity. For example, Wills et al. (2013) placed adults under working memory load by having them perform a secondary delayed recall task and showed that the likelihood of unidimensional (i.e., child-like) categorization increases under working memory load, resulting in broader categories.

There has been little work examining developmental differences in categorizing temporally extended patterns. The few studies that exist suggest that children are less likely than adults to attend to features whose detection requires temporal integration. In particular, Schwarzer (1997) found that while adults were more likely than children to focus on attributes like tempo and contour, whose perception requires some degree of temporal integration, children (5- to 7-year-olds) tended to focus on loudness and timbre, which could be assessed at any point during a melody in Schwarzer’s stimuli. As described above, the phonological literature on intonation suggests that a contour for adults is defined by a temporally extended sequence of necessary features, which requires temporal integration for detection. Schwarzer's findings then predict that children are more likely to have broader intonation contour categories than adults.

In addition to resulting in broader categories, the proposal that intonation contour perception poses a working memory challenge also argues against veridical encoding of experienced contour exemplars. Veridical encoding of exemplar details is necessary for observing a categorization advantage for old exemplars over comparable new exemplars (e.g. Nosofsky, 2000; J. Smith & Minda, 2000). Therefore, reduced working memory is expected to diminish or eliminate the advantage that experienced exemplars are often observed to have over novel exemplars in studies of categorization (see Goldinger, 1998; Port, 2007; J. Smith & Minda, 2000, for reviews). In particular, while individual word exemplars may be stored with detailed pitch contour information (Schweitzer et al., 2015; Walsch et al., 2013), this appears less likely for temporally extended contours.

If adults *are* able to veridically encode contour exemplars and categorize based on analogy to stored exemplars, children's reduced ability to veridically store exemplars should again lead to broader categories. This prediction is suggested by the results of Gibson & Gibson (1955), who reported an experiment in which 6- to 11-year-old children and adults were repeatedly presented with a single scribble and then were asked to judge other scribbles differing from it in various ways. The task was to decide whether the new scribble was or was not a member of the same category as the training scribble. The results were a developmental shift from "everything is in the category" to "only experienced examples are in". However, studies on categorization of stimuli with multiple separable dimensions have suggested that both children and adults attend to a subset of the stimulus features, although adults are more likely to attend to more than one feature (Thompson, 1994; Ward et al., 1990; Ward & Scott, 1987). Thus, it is also entirely possible that intonation contours are too complex for either children or adults to encode veridically, resulting in no advantage for old exemplars over comparable new exemplars in either subject group.

1.3. Current study

Our study focuses on intonation contour learning in 9- to 11-year-old children and adults. We expected that category learning biases favoring broad categories would diminish during this age range based on similar findings in developmental studies of non-linguistic categorization (e.g. Berger & Hatwell, 1996; Gibson & Gibson, 1955; Schwarzer, 1997), and because working memory continues to develop into adolescence (Luna et al., 2004). In addition, it has been suggested that late childhood (8 to 11 years) is an important transition time for the ability to acquire complex sound patterns. In particular, Tahta et al. (1981) examined the ability of children aged 5 to 15 years to replicate foreign intonation and found that the “ability to replicate intonation remained steadily good until 8, then dropped rapidly until 11.” (ibid, p.363). More generally, Labov (2010, p.8) argues that “9-10 appears to be a critical age for entering a new community” in that children who enter a new community later tend not to acquire the vowel categories of that community and the details of their phonological conditioning. Like Tahta et al. (1981), he argues that the language learning mechanism becomes adultlike by 11.

Perception of speech in noise continues to improve until age 12 (e.g. Elliott, 1979; Hartley et al., 2000; Schneider et al., 1989; Siegenthaler, 1969). Improvements are also seen until around this age in non-linguistic auditory discrimination tasks. 9-10 year-old children are found to have difficulty detecting a tone before an immediately following masking noise, suggesting difficulties with temporal resolution (Buss et al., 2013; Fox et al., 2012) or processing efficiency (Hartley et al., 2000; Hill et al., 2004). Moore et al. (2011) report data from a variety of psychophysical auditory discrimination tasks in a large sample of children aged 6 to 11, suggesting that "maturity was achieved... by 10 to 11 years of age on all tests" (ibid, p.269) except even 11-year-olds were not completely adultlike at being able to detect which tone, in a sequence of three, is higher than the others. The continued improvement in speech perception and sound discrimination may be linked to the fact that, while the peripheral auditory system matures quite early, structural maturation of the human auditory cortex is not complete until age 11 or 12 (Moore & Guan, 2001; Moore & Linthicum, 2007).

In this study, we examined 9-11 year-olds. Based on the research reviewed above, we suspected that this age group may let us capture the end of the developmental trajectory, so that by the end of this period children would be adultlike in their auditory category learning biases.

In order to examine these biases, we used a modified version of the schema abstraction paradigm (Posner & Keele, 1968). The schema abstraction paradigm is a classic approach to the study of category learning, used extensively in visual categorization, which was designed to investigate the acquisition of novel categories of perceptual patterns defined within a continuous similarity space. Child and adult listeners were trained to associate contour exemplars with one of three alien languages. As in Posner & Keele (1968), the exemplars consisted of low-level distortions of category prototypes which were themselves withheld during training. At test, listeners were presented with trained exemplars, prototypes, new low-, mid- and high-level distortions, and, for one category, an additional set of distractors. The test task was to either classify each contour into one of the trained categories, or to reject it as belonging to none of the trained categories. This 'none of the above' option was unavailable in Posner & Keele (1968) and lets us shed additional light on possible age-related changes in category breadth.

Descriptions of intonation contours like H L* H* L% indeed suggest that the task of intonation contour category learning entails schema abstraction, as the categories are defined by examining the contours associated with a particular meaning and identifying what these contours share. As a result, the intonation contour category is defined by the full set of features present in its exemplars, rather than by the features that are most informative in distinguishing the category from other categories. This is based in the fact that intonation contours have to be produced as well as perceived and the belief that the schema is shared between production and perception. While a listener can often succeed in discriminating a contour from others by paying attention to a small subset of the contour's features, no features can be skipped in production (even ones shared with other contour categories). If intonation contour learning is to subserve production (and any aspect of language learning probably does) the learner needs to identify what contours with a certain meaning are like, resulting in a schema that can be used to guide re-generation the contours by providing a perceptual target that the production system must learn to generate. This is quite different from perceptual category learning in domains where the learned category is not intended to provide a production target, such as visual pattern learning, where accurate discrimination between categories, and hence acquisition of a discriminative system of cues, is all that is required (e.g. Nosofsky, 2000). The focus on schema abstraction motivates our decision to present intonation contours blocked by category (Carvalho & Goldstone, 2012, 2014), and our concern with acceptability rather than confusability, as described below.

2. Methods

2.1. Participants

Forty children and forty adults participated in the experiment. The adults were recruited from introductory psychology classes through the University of Oregon Psychology/Linguistics Human Subject Pool. They received course credit for participating. The children were 9 to 11 years old and were recruited from local schools, by word of mouth and through a database maintained by the University of Oregon's Psychology Department. There were sixteen 9-year-olds, fifteen 10-year-olds, and nine 11-year-olds (median age10;3). The age distribution for the children was close to uniform (D=.11, p=.89 according to the Kolmogorov-Smirnov test), providing a good sampling of ages within the range. The children received a small gift for participating, and families were compensated for their time.

2.2. Stimuli

The stimuli consisted of intonation contours superimposed on a 15-syllable nonsense carrier phrase 3.43 seconds in length. The carrier was resynthesized from a recording of a sentence uttered by a male speaker of U.S. English obtained from the online Speech Accent Archive (Weinberger, 2013). Resynthesis consisted of replacing every consonant of the utterance with /m/ and every vowel with /i/, followed by diphone concatenation in MBROLA (Dutoit et al., 1996). The resulting nonsense "mimimi" string thus preserved the rhythm characteristics of spoken English.

The contours were derived from one of three prototypes that constituted category centers. The prototypes were designed to vary in complexity (number of characteristic features). The simplest, 'flat' prototype featured a constant F0 of 250 Hz throughout the carrier duration. The 'final fall' prototype proceeded at 250 Hz until the onset of the 11th vowel (V11), falling to 110 Hz over the final 4 syllables (a 14-semitone drop). The multi-feature M prototype rose from 250 Hz to 350 Hz at V4 (6 semitones), dipped to 215 Hz at V6 (8 semitones), held until V11, and rose again to 340 Hz at V13 (8 semitones) before falling to 175 Hz (11 semitones) for the final syllable.

The M prototype was designed, in part, to begin studying the relative importance of the beginning and end of an intonation contour, i.e., to investigate feature weighting. For this reason, the peaks of the M prototype were designed to be perceptually equal so that if one of the peaks were found to matter more for category membership, the effect could not be ascribed to perceptual salience but instead to feature weighting within the category representation. Given that F0 usually declines throughout an utterance (e.g. Cohen & ‘t Hart, 1967; Ladd, 1984), listeners expect declination and perceptually compensate for this (e.g. Pierrehumbert, 1979; Gussenhoven et al., 1997). We designed the stimuli to accommodate perceptual compensation by making the second peak lower than the first peak by 0.5 semitones. To ensure that this adjustment was of an appropriate magnitude, 30 native speakers of English were asked to provide binary judgments of relative peak height (as in Gussenhoven et al., 1997) for various examples of the M contour. The height of the first peak was fixed to be equal to that of the original M prototype, while the height of the second peak was raised or lowered in .25-semitone increments up to 2 semitones distance from the original position. A logistic regression model predicting response from F0 difference identified the second peak height 0.08 semitones above that of the second peak height in the M prototype as the threshold below which the first peak is perceived to be higher than the second peak. This is in line with the just-noticeable-difference for a pitch height difference between tones separated by 2 seconds (as are the peaks in our stimuli) from a meta-analysis of psychoacoustic studies on pitch change judgments reported in ‘t Hart et al. (1990, p.32). Thus compensation for declination was properly controlled in the M prototype: the two peaks appear to be perceived as approximately equally high without training, at least for adults.

Training items were created in Praat (Boersma & Weenink, 2009) by randomly perturbing the prototypes such that each inflection point and vowel midpoint had an equal chance of rising by one semitone, falling by one semitone, or remaining 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 were made for a total of twelve training exemplars per category.

Test items included four trained exemplars, as well as four novel low-level distortions created from the prototypes in identical fashion. They also included the prototypes themselves, as well as 4 novel mid- and 4 high-level distortions wherein the pitch of each syllable of the prototype was perturbed by up to 3 or 5 semitones, respectively. These random distortions of the prototype were intended to investigate category breadth and veridical encoding of experienced exemplars. A maximally narrow category would include only the experienced exemplars, so that even novel exemplars that are as far from the prototype as experienced exemplars are rejected from the category. Learning such narrow categories requires veridical encoding of acoustic detail so that experienced exemplars can be distinguished from similar novel distortions. A somewhat broader category could include novel distortions that are as far from the prototype as experienced exemplars. An even broader category would include new mid-level distortions of the prototype, which are farther away from the prototype than the experienced exemplars. A very broad category would also include high-level distortions.

Finally, three sets of distractors for the M category were generated by removing either the first peak, the second peak, or the valley from the prototype and creating two pairs of low-level distortions from the result. These stimuli were designed to both investigate category breadth and the importance of various characteristic features of a contour. A broad category would include distortions lacking some of the features shared by all of the experienced exemplars. For example, test stimuli lacking the first peak or the second peak of the M prototype might still be considered instances of M. High importance of a feature for categorization would result in stimuli lacking that feature not being assigned to the same category as experienced exemplars possessing the feature. We particularly focused on the two peaks of the M prototype as features or complexes of features characterizing the M contour. Recall that the two peaks were designed to be perceptually equal. As a result, responses to test stimuli lacking one of the peaks provides information on the feature weights associated with the early peak vs. the late peak for categorization of intonation contours. L. Smith (1989) suggested that adults may be more likely to assign different weights to features characterizing a category (cf. also Hill et al., 2004, p.1027; Oh et al., 2001). If found, this pattern would manifest itself as one of the peaks mattering more than the other for adult categorization, with the two peaks mattering relatively equally for children.

Example training and test items are shown in Fig. 2. All stimuli are available online at https://www.dropbox.com/sh/qs9uem17i4t8ri2/AAB5ROSoVHkyTZQFkSjUjJ0na?dl=0

FIGURE 2 ABOUT HERE

2.3. Procedure

The experiment was administered in a quiet room, with the participant seated in front of a computer screen while wearing headphones. The training stage consisted of auditory presentation of contour exemplars paired with pictures of alien creatures who "talk like this." There were three such creatures, each of which was always paired with exemplars from the same set ('flat', 'final fall' or M).

Learning was passive, with items advancing automatically after an interval of 500 milliseconds. Following Carvalho & Goldstone (2012, 2014), training was blocked by category in order to encourage participants to pay attention to features that members of the same category have in common as opposed to attending to contrastive features, i.e., those that distinguish the categories. This design choice was motivated by our belief that intonation contour categories are defined by all of the features that members of the category share, rather than the features that are useful to distinguish the contour category in question from other categories. Each block consisted of the 12 training items presented in random order. Block order was likewise randomized, and each block was presented 3 times. The training lasted approximately 9 minutes.

The test stage immediately followed training. Test items were presented auditorily to participants over headphones in random order, and participants categorized the heard contours by clicking on one of four creature options available on the screen. Three of the choices consisted of pictures of the familiar alien creatures. The fourth was a picture of a group of several new aliens (Fig. 3). The participants were asked if the "sentence" they heard “was said by a *creature 1*, a *creature 2*, a *creature 3*, or one of these other aliens?". Item presentation advanced 500 milliseconds after registering a click on one of the pictures. There were a total of 72 test items (4 exemplars x 5 distortions x 3 categories + 12 M distractors). Testing lasted approximately 6 minutes.

FIGURE 3 ABOUT HERE

2.4. Analysis

Statistical analyses were performed using mixed-effect logistic regression as implemented in the lme4 package (Bates & Maeschler, 2010) in R (R Development Core Team, 2010). The dependent variables were acceptance (was the ‘correct’ category or the ‘one of these other aliens’ response selected) and confusion (was the ‘correct’ category or some other experienced category selected). The data on acceptance thus excludes confusions, and the data on confusion excludes clicks on ‘one of these other aliens’. Both dependent variables are binary trial-level variables. We therefore used participants and items as crossed random effects and included within-participant random slopes for within-participant predictors: category and distortion level.[1] We included within-item random slopes for the between-participant predictors: age and subject group (child vs. adult), resulting in maximal random-effects structure (following Barr et al., 2013).

Based on both the design of the stimuli and our primary interest in category breadth, we focused on acceptance rather than confusion probability since increased category breadth is most purely indexed by increased acceptability (reduced probability of using the ‘none of the above’ option) insofar as.greater acceptability of low-confusability exemplars is difficult to account for by slower or poorer learning. Broad categories may also lead to increased confusability of the trained categories; however, because we conceive of intonation contour category learning as being about learning the characteristics of contours sharing a meaning rather than about discriminating between contour categories, unacceptable test stimuli were not designed to be more confusable than acceptable ones. While distractor and high-level prototype distortions were made to be less similar to the prototype than the training examples, they were not made to be more similar to the prototypes of the other categories; that is, the distractors and high-level distortions were not restricted to deviate from their prototype in the direction of other prototypes. As one might expect, given this stimulus design, there were few significant differences in confusability between the various stimulus types within any subject group. Nonetheless, confusion data analyses are reported in footnotes. Full confusion matrices are presented in the Appendix.

---

[1] Recall that each category has one unique prototype. Items at different distortion levels were created independently of each other by randomly perturbing the prototype for the category, rather than creating higher-level distortions from lower-level distortions. Distortion level is therefore a between-item variable.

We operationalized children’s age in one of two ways: continuous age in months and a binary split at the median (10;3). The results are unchanged across these two operationalizations. Following injunctions against dichotomization (e.g. MacCallum et al., 2002), we used continuous age in regression analyses investigating the effect of age within the child group. Binary age is used *only* for the figures because the few data points per response category per age rendered scatterplots uninformative, . Error bars in the graphs are 95% confidence intervals based on a proportion test (prop.test() in R, Wilson, 1927).

3. Results and Discussion

For each type of test stimulus, we report three statistical analyses: one comparing children to adults (group), one examining the effect of age within the child group (children’s age), and one examining whether older children differ from adults, with child age dichotomized via a median split. In all statistical analyses reported, the dependent variables are binary (confusion and acceptance, as defined above); however, full stimulus category by response category by participant group cross-tabulations are presented in the Appendix.

3.1. Old and new low-level distortions

We began by examining possible effects of group and children’s age on responses to old and new low-level distortions. We expected the intonation contours to pose a working memory challenge for both groups, enough to prevent reliance on veridical memories of specific exemplars for categorization, thus predicting a null result for this analysis.

As shown in Fig. 4 below, neither adults nor children showed a significant difference in responses to old and new exemplars one semitone away from the prototype (b = -0.14, $se(b)$ =

0.32, *z* = -0.43, *p* = .67 for adults; *b* = -0.13, *se(b)* = 0.35, *z* = -.38, *p* = .71 for children); that is, there was no exemplar effect for either group. There was no significant interaction between subject group and stimulus type (*b* = 0.06762, *se(b)* = 0.29233, *z* = 0.231, *p* = 0.817), nor between age and stimulus type within the child group (*b* = -0.02532, *se(b)* = 0.20142, *z* = -0.126, *p* = 0.900).[2]

FIGURE 4 ABOUT HERE

While null results should be interpreted with caution, it is worth pointing out that the difference between old and new exemplars tends to be very robust in simple visual categorization experiments. For example, J. Smith & Minda (2000, p.11, Fig. 3) showed that a minimal change to an old exemplar makes the changed stimulus 87% more dissimilar from the old exemplar in Nosofsky's (1986) exemplar-based Generalized Context Model of classic visual categorization data, leaving only 13% for the variance among novel exemplars. Thus, a difference in acceptability between old and new exemplars should be expected if they are being veridically remembered and relied upon for making the reject/accept decision. As we see in Fig. 4, though, distance from the prototype among novel exemplars appeared to matter much more than old/new status in the present experiment. Thus, intonational contour exemplars are either not being veridically remembered or are not being relied upon for categorization, or both.

---

[2] Children show higher confusion rates than adults (*b* = -0.94, *se(b)* = 0.41, *z* = -2.30, *p* = 0.021), and confusion decreases with age (*b* = 0.57, *se(b)* = 0.24, *z* = 2.42, p = 0.016), but there are no differences in confusability between new and old exemplars within either age group: (*b* = 0.37, *se(b)* = 0.46, *z* = 0.81, *p* = 0.42 for adults; *b* = -0.46, *se(b)* = 0.40, *z* = -1.16, *p* = 0.24 for children). There is no significant interaction between subject group and stimulus type (*b* = -0.29, *se(b)* = 0.39, *z* = -0.75, *p* = 0.45).

3.2. Random high-level distortions

Categorization of higher-level (3 semitone and 5 semitone) distortions of the prototypes provide one of the most direct tests of category breadth. If children's categories are broader than those of adults because they retain fewer tonal features than adults, children should show a smaller difference between low-level and high-level distortions. Young children should show an even smaller difference between distortion levels than older children since feature retention (and integration) is dependent on working memory capacity, which changes continuously with age (e.g. Luna et al., 2004).

The results are summarized visually in Fig. 4 above, with the continuous effect of age represented categorically using a median split.[3] There is a clear effect of distance from the prototype, with larger deviations rejected more often than smaller deviations (main effect of distortion level, Table 1), and the effect of distance from prototype increases with age (see Table 1 for children vs. adults, and Table 2 for continuous effects of age within the child group). Most strikingly, the youngest children accepted 3-semitone distortions as often as 1-semitone distortions. Fig. 4 also shows that the age range we examined likely shows the endpoint of the developmental trajectory: in agreement with Tahta et al.'s (1981) repetition data, the older half of the children appear adult-like in their behavior in Fig. 4 and are not in fact statistically different from adults (Table 3).[4]

---

[3] Note that we are not claiming that there is a developmental transition at 10;3. The decreases in acceptability appear to be fairly continuous throughout the examined age range.

[4] Children show more confusion than adults ($b = -1.09$, $se(b) = 0.41$, $z = -2.67$, $p = 0.008$), and confusion decreases with increasing age within the child group ($b = 0.85$, $se(b) = 0.26$, $z = 3.24$, $p = 0.001$). However, while 3-semitone and 1-semitone distortions differ in acceptability ($b = -2.07$, $se(b) = 0.26$, $z = -8.11$, $p < .0001$ for adults; $b = -1.31$, $se(b) = 0.29$, $z = -4.48$, $p < .0001$ for children), they do not differ in confusability for either subject group ($b = -0.40$, $se(b) = 0.43$, $z = -0.93$, $p = .35$ for adults; $b = -0.08$, $se(b) = 0.32$, $z = -0.27$, $p =.79$ for children). 5-semitone distortions are significantly more confusable than 1-semitone distortions for both children

TABLE 1 ABOUT HERE

TABLE 2 ABOUT HERE

TABLE 3 ABOUT HERE[5]

3.3. Feature-removing distortions

We now turn to categorization of the M distractors. These distractors might be seen as high-level distortions of the M prototype, but, unlike the distortions examined in the preceding section, they were not random distortions. Instead, each distractor eliminated one of the features characteristic of all M training exemplars and the M prototype. This manipulation provided us with another way to test whether children's categories are broader or narrower than those of adults.

Fig. 5 provides a graphical illustration of the results, while Tables 4 and 5 report the results of mixed-effects regression modeling. Table 4 suggests that children were more likely to accept the early peak distractors as instances of the M contour than adults. However, Table 6 shows that this is really driven by the difference in behavior between younger children and the rest of the participants. Thus, the interaction between distractor type and subject group in Table 4 is not statistically reliable according to an omnibus test, suggesting that children as a group

---

and adults (b = -0.73, $se(b)$ = 0.32, z = -2.29, p = 0.022, b = -1.01, $se(b)$ = 0.46, z = -2.20, p = 0.028 respectively). However, there are no significant interactions between distance (1 vs. 5) and either age or subject group (b = 0.37, $se(b)$ = 0.39, z = 0.93, p = .35, b = -0.08, $se(b)$ = 0.24, z = -0.326, p = .74 respectively) with confusion as the dependent variable.

[5] There was also a significant interaction between distortion level and stimulus category. However, we did not include interactions with stimulus category (flat, final fall, M) in the presented model because the observed difference between 1-, 3- and 5-semitone distortions holds within each stimulus category. Although it is significantly smaller in magnitude for the category formed from the flat prototype, 3-semitone and 5-semitone distortions differ from 1-semitone distortions even for that subset of stimuli (b = -1.63, $se(b)$ = 0.38, z = -4.30, p < .0001 for 1 st vs. 3 st; -b = -3.10, $se(b)$ = 0.44, z = -7.01, p < .0001 for 1 st vs. 5 st).

may not differ from adults in acceptance of deviations from M. In contrast, Tables 5 and 6 show that there is a significant age-related change within the child group, the difference in acceptance rates for distractors and novel exemplars of M growing with age. These results are again consistent with the prediction that younger children's categories are broader than adults' categories but children's categorization becomes adult-like by the end of the period under investigation.[6]

FIGURE 5 ABOUT HERE

TABLE 4 ABOUT HERE

TABLE 5 ABOUT HERE

TABLE 6 ABOUT HERE

Although adults and older children rejected the early peak distractor as instances of the M contour, acceptance rates were higher for late peak distractors: ($b = 2.05$, $se(b) = 0.36$, $z = 5.73$, $p < .0001$ for late peak vs. early peak; $b = -2.87$, $se(b) = 0.42$, $z = -6.78$, $p < .0001$ for late peak vs. hat). Hat distractors were rejected slightly more often than early peak distractors ($b = -0.80$, $se(b) = 0.38$, $z = -2.10$, $p = 0.036$). Importantly, though, late peak distractors were accepted significantly less often than novel instances of the M contour by adults ($b = -3.98$, $se(b) = 0.86$, $z = -4.60$, $p < .0001$), suggesting that adults did not attend exclusively to the second peak in making category judgments.

---

[6] The age-related decline in distractor acceptance within the child group is especially noticeable for distractors retaining only the first peak of M in Figure 5. Whereas adults and older children categorically rejected such examples, younger children frequently accepted them as instances of M. However, within the group of distractor stimuli, the interaction between distractor type and age is not significant in an omnibus test ($\chi^2(2) = 2.25$, $p = .32$).

In contrast to adults and older children, younger children were equally likely to accept distractors that retained either the first or the second peak of the M prototype. They were also as likely to accept single-peaked distractors as novel two-peaked instances of M contours that were one semitone away from the prototype (*b* = -0.86, *se(b)* = 0.65, *z* = -1.33, *p* = .18 for M vs. early peak; *b* = -0.57, *sd(b)* = 0.78, *z* = -0.74, *p* = 0.46 for M vs. late peak; *b* = 0.08, *se(b)* = 0.54, *z* = 0.15, *p* = .88 for early peak vs. late peak). As with adults and older children, the hat distractor was still more likely to be rejected than the other distractors (*b* = -2.67, *sd(b)* = 0.83, *z* = -3.24, *p* = 0.001 for M vs. hat). Apparently, the level high pitch of the hat distractor was experienced very differently than the peaky M contour.[7]

In all, then, the results on the categorization of M distractors indicate that children's categories are broader than adults', consistent with the predictions based on working memory differences. A distractor retaining either peak of the prototype was judged by younger children to be an instance of the M contour category, consistent with the idea of a lower dimensional schematic representation of the contour as "peaked." By contrast, adults and older children only accepted distractors retaining the late peak as members of the M contour category, indicating that their category was more specific than younger children's categories, in that the acquired schema requires both peaks to be present. Given the perceptual equivalence of the two

---

[7] There was no significant difference in confusion between children and adults (*b* = -0.64, *se(b)* = 0.56, *z* = -1.14, *p* = .25), although confusion decreased as age increased within the child group (*b* = 1.04, *se(b)* = 0.42, *z* = 2.48, *p* = .013). There were few significant differences in confusability between the stimulus types for any subject group. For adults, none of the distractor types differed from M: *b* = -0.26, *se(b)* = 0.39, *z* = -0.67, *p* = .50 for early peak vs. M; *b* = -0.5187, *se(b)* = 0.32, *z* = -1.64, *p* = .10 for late peak vs. M; *b* = -0.28, *se(b)* = 0.42, *z* = -0.656, *p* = .51 for 'hat' vs. M. For children, early peak and late peak stimuli did not differ from M in confusability (*b* = -0.49, *se(b)* = 0.45, *z* = -1.10, *p* = .27; *b* = -0.54, *se(b)* = 0.41, *z* = -1.31, *p* =.19 respectively) but the 'hat' was somewhat more confusable (*b* = -1.19, *se(b)* =0.57, *z* = -2.10, *p* = 0.04). There was no significant effect of subject group (child vs. adult) on confusability of 'hat' (*b* = -0.65, *se(b)* = 0.71, *z* = -0.91, *p* = 0.36) and a marginal effect of age within the child group in the unexpected direction, with 'hat' becoming more confusable with age (*b* = -0.99, *se(b)* = 0.52, *z* = -1.90, *p* = .06).

peaks for adults (as documented in the methods section), the greater importance of the second peak for adults and older children was unexpected.

4. General Discussion

The present study extends research on category learning to the acquisition of intonation contours, temporally distributed fluctuations in a continuous parameter, namely, pitch. We hypothesized that the great and greatly variable extent of intonation contours poses challenges to working memory that should result in younger children acquiring broader categories defined by fewer necessary features compared to older children and adults. The results described above were consistent with this prediction.

4.1. Working memory and category breadth

Why are children so open-minded? The hypothesis that motivated the present study links this open-mindedness to low working memory capacity in particular, and to children's cognitive limitations more generally. Limitations of working memory may affect category breadth directly or indirectly by reducing children's processing speed (e.g. Ito et al., 2014, for processing intonation; Buss et al., 2013, and Fox et al., 2012, for a possible role of slower processing in the difficulties faced by children in auditory masking tasks and non-word repetition respectively, but cf. Hall et al., 2004). If categorizing on the basis of multiple dimensions is slower and more resource-demanding than categorizing on the basis of a single dimension (Berger & Hatwell, 1996; Ward & Scott, 1987; Wills et al., 2013), and children’s working memory and processing speed are limited, then children may prefer to respond on the basis of a single dimension to reduce effort.

A related possibility is that children's *perceptual* abilities are underdeveloped compared to those of adults (Gibson & Gibson, 1955; Hill et al., 2004; Moore et al., 2011; Odom & Cook, 1984; Thompson & Massaro, 1989), resulting in blurrier representations of the individual feature values. In particular, even 11-year-old children were found to be poorer than adults at discriminating tones in an AXB task (Moore et al., 2011), and to be more susceptible to interference from auditory maskers following the to-be-identified target tone (Buss et al., 2013). Buss et al. (2013) suggest that children are unable to separate out auditory stimuli that closely follow each other in time, averaging them together. Hill et al. (2004) and Werner & Boike (2001) propose that children have a higher level of internal noise, which compromises veridical encoding and the veracity of encoded auditory signals. The resulting blurriness of the individual pitch values forming a contour may lead to broader categories. Under this hypothesis, children may not only represent fewer features per contour example but also have blurrier representations of values of the features they do represent. This hypothesis would help explain why children are more likely to accept *feature-preserving* deviations from the prototype than adults are (particularly, 3-semitone distortions in Fig. 4).

It may also be the case that children are generally more accepting than adults at the decisional level (cf. Hall et al., 2004). Knowing that their memory is limited, children may be more accepting of mismatches between current perception and memory. Similarly, willingness to accept deviations from prior experience may be expected of participants with limited prior experience (Xu & Tenenbaum, 2007). This may be functional in that it allows child learners to be more open to changing their prior beliefs (e.g. Gopnik et al., 2015). It may also be useful to cover the perceptual space with one's limited set of categories: when most stimuli are novel, it is helpful to be able to assimilate these novel stimuli into existing categories to be able to venture a guess as to their behavior (Rogers & McClelland, 2004).

The greater breadth of children's categories is in line with much previous work on children’s categorization (e.g. Gibson & Gibson, 1955; Lawson & Fisher, 2011; Mandler, 2000; Pauen, 2002). It is also consistent with developmental work on word recognition, which can be interpreted as examining the breadth of categories delimiting possible orthographic or acoustic realizations of the same word. Thus, in spoken word recognition, young children have been found to be remarkably tolerant of mispronunciations of unfamiliar words (Fennell & Werker, 2003; Swingley, 2007). In visual word recognition, Castles et al. (2007) showed substantial priming between minimally different spellings, e.g. *lpay* → *play*, in 3rd graders that disappeared by 5th grade, suggesting that *lpay* is assimilated into the *play* category by younger children. In syntax, children have been argued to subsume a broader range of sentences under a single construction. For example, Rowland et al. (2012) found that verb overlap between prime and target increased the amount of priming for adults and older children but not younger children, who exhibited more priming of abstract syntactic patterns independently of lexical overlap. In phonology, Cristia & Peperkamp (2012) have argued that infants form broader generalizations about possible word onsets than adults do.

In spite of the wide variety of reasons to explain why children may form broader categories than adults, there are also many studies in which children were documented to form narrower generalizations than adults or where younger children were found to be more narrow-minded than older children (e.g. Rabi & Minda, 2014b, for visual patterns, Boyd & Goldberg, 2012, and Savage et al., 2003, for syntactic constructions, Jenkins et al., 2015, for semantic categories). This suggests that children are not always more broad-minded than adults, and raises the question of *when* they broad-mindedness is more likely to be observed. We suggest that children are especially likely to form broader categories than adults when narrow generalization would pose too high a load on working memory or other processing resources. This suggestion is consistent with findings on the relationship between working memory and category formation. For example, Wills et al. (2013) observed that adults were less likely to pay attention to multiple features characterizing a category when placed under working memory load, resulting in broader categories. In addition, Spencer et al. (2011) and Lawson & Fisher (2011) found that sequential presentation of category exemplars, which may place higher demands on working memory than untimed simultaneous presentation, encouraged broader generalization in children while having a much smaller influence on adults (see also Lawson, 2014).

4.2. Representations of intonation contour categories

It appears that neither children nor adults relied on veridical memory for specific exemplars to determine category membership. 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 Figure 4. This result contrasts with classic findings for simple visual patterns where the difference between old and new exemplars is typically extremely robust to the extent that all novel exemplars are more similar to each other than to familiar exemplars (J. Smith & Minda, 2000).[8] The lack of a difference between old and new exemplars suggests a relatively abstract,

---

[8] Nosofsky (2000) objects to this conclusion of J. Smith & Minda’s (2000), pointing out that novel exemplars and familiar exemplars may be very similar in how far they are from experienced exemplars of other categories. Thus, despite having low similarity to each other, old and new exemplars may nonetheless have similar probabilities of being confused with members of other categories. However, this objection does not apply to our acceptance data, since no examples of the ‘none of the above’ category were presented in training.

or underspecified, representation of intonation patterns relative to visual patterns, abstract enough for the new exemplars not to differ from the old exemplars.

There are many ways in which intonation contour exemplars differ from the exemplars of simple visual patterns that are the mainstay of work on categorization (J. Smith & Minda, 2000). First, they are undeniably more multidimensional: in our stimuli, every syllable could vary in pitch by various amounts for a total of 16 degrees of freedom. In contrast, the most complex stimuli we have seen in the visual categorization literature are the 8-dimensional stimuli seen to disfavor exemplar memorization in Minda & J. Smith (2001). Contours, unlike the visual patterns, are also extended in time and thus require temporal integration in working memory (see also Schwarzer, 1997). Difficulties with integration may make learners more likely to focus on a limited set of attributes, not even detecting a difference between old and new exemplars in the unattended details (e.g. Wright et al., 2000; Olsson & Poom, 2005). Future work should consider whether veridical storage of acoustic details of intonation contours occurs when the contours in question is relatively short, as suggested by Schweitzer and colleagues' work on pitch accents (e.g. Schweitzer et al., 2015).

The structure for natural intonation contour categories, as well as other meaningful categories of linguistic forms, differs from that examined in most classic studies of categorization showing exemplar memory effects. Most studies of "natural categories" have examined ill-defined categories (e.g. McCloskey & Glucksberg, 1978; Rosch & Mervis, 1975), where no features are individually necessary and jointly sufficient for category membership. However, natural categories of phonological forms *are* well-defined (Kapatsinski, 2014, Stockall & Marantz, 2006). It is for this reason that, in the absence of noise, mismatch in a single feature or letter is sufficient to eliminate repetition priming. For example, the pseudoword *blick* -- if correctly perceived -- does not prime the recognition of *brick* any more than the pseudoword *nabe* would (Castles et al., 2007; Glezer et al., 2009), suggesting that all letters or phonological features need to be perceived as having been produced by the speaker/writer for the word to be recognized. Similarly, ‘t Hart et al. (1990, p.84) describes intonation contours as characterized by an invariant set of melodic properties instantiated in many different pitch contours. With such a category structure, there may be less reliance on memorizing individual exemplars than there would be with an ill-defined category (Blair & Homa, 2003; Reed, 1978).

Exemplar models represent each exemplar as an island within the archipelago of a category: the shape of the category itself is unconstrained (e.g. McKinley & Nosofsky, 1995), but there is a major difference between old exemplars and new exemplars just as there is a difference between being on an island vs. in the water between islands. In contrast, meaningful categories of linguistic forms, like the set of instances of an idiomatic intonation contour, may be better represented as single islands: as long as a novel stimulus is within the island, it is part of the category, with distance from the island reducing similarity to the category in a continuous fashion. Language learners may come to know that categories of phonological forms tend to be structured this way, and may therefore expect it even of novel contour categories. It remains to be seen whether very young language learners also acquire language categories the same way as the children examined in this study. The idea that they might not follows from the hypothesis that it is language experience that explains why our listeners did not remember, or categorize on the basis of, contour exemplars.

It could, however, be that we did not find a reliance on exemplars in test because unlike most studies of categorization, including the classic Posner and Keele (1968) study, we did not train participants to criterion on categorizing the training exemplars. Training and testing on the same exemplars to criterion may lead study participants to focus on remembering the specific exemplars presented, thus reducing generalization to novel exemplars. Indeed, J. Smith & Minda (1998) provide evidence for reduced exemplar memory effects in the early stages of training. In the present study, participants received a fixed amount of training. Furthermore, the amount of training was fairly limited. Perhaps, more extensive training is needed to form lasting veridical memories of specific exemplars of intonation contours. For phonological units, gradual development of memory for specific experienced structures accompanied by rapid sensitivity to broad generalizations is shown by Linzen & Gallagher (2014).

4.3. Final prominence

A final, unexpected finding was that the late peak was more important than the early peak for M category membership for adults and older children whereas younger children appeared to weight both features equally. A simple perceptual explanation for this result is ruled out: the two peaks were judged approximately equally high without training, thus their absence should have been equally salient.

A developmental trajectory from (near-)equal weighting of features to selective weighting of features has been proposed for object recognition in vision (L. Smith, 1989) as well as for speech sound perception (Pisoni et al., 1994). Our results provide support for this hypothesis. However, note that L. Smith (1989) proposed that children perceive stimuli more holistically due to exercising less selective attention. Our results do not agree with this aspect of Smith's theory: while older children and adults weighted the second peak more than the first peak, they also judged stimuli lacking both peaks to be instances of M less often than stimuli that have both peaks. In contrast, younger children weighted the two peaks equally but accepted stimuli having either peak as instances of M as often as they accepted novel one-semitone deviations from the prototype or even previously experienced exemplars.

Why is it the second peak that is singled out as particularly important by older children and adults? We suggest that older children and adults preferentially attend to pitch peaks located close to the end of the utterance because it is usually the most informative part, hosting new information (Clark & Clark, 1978; Wundt, 1900), and is therefore likely to contain pitch peaks cueing focus (e.g. Choudhury & Kaiser, 2012). The developmental trajectory we observed fits with Narasimhan & Dimroth's (2008) conclusion that 3- to 5-year-old children, having not yet learned that new information comes at the end of the sentence, tend to place new information at the beginning. The importance of the second peak for older children and adults also fits well with natural-language results of ‘t Hart et al. (1990) and Choudhury & Kaiser (2012). ‘T Hart et al. (1990, p.88) write that intonation “contours derive their ‘pattern identity’ … not [from] the last pitch movement but [from] some (near) final melodic structure as a whole”, in our case this near-final melodic structure being the second peak. Choudhury & Kaiser (2012) found that the height of the second peak in natural two-peaked contours in Hindi and Bangla varies with focus type, while the height of the first peak does not. They suggest that “prosodic distinctions between focus types are amplified at the default focus position” (p.4), which is generally close to the end of the utterance. However, it is also possible that adults paid more attention to the end of the contour in M because it is the most informative part of our set of contours: attending to the end pays off for distinguishing the flat prototype and the final fall prototype, which differed only in their final portion.

5. Conclusion

The present study extends work on perceptual learning to temporally distributed, pitch patterns, namely intonation contours. Both children and adults were found to rapidly learn new intonation contour categories. Both groups formed relatively abstract representations of contours, with no advantage for familiar exemplars over new exemplars as long as the new and familiar exemplars were equidistant from the prototype. However, the categories they formed did differ. Given the same experience, 9- to 10-year-old children formed broader categories than 10- to 11-year-old children and adults in that they were more likely to accept major distortions of the prototype into the category. This broad-mindedness included even distortions that removed some features that were present in all training examples. In addition, older children and adults were found to pay more attention to the end than to the beginning of the contour for the purposes of categorization.

The paradigm we used to study contour learning opens a number of avenues for future research. First, students of intonation may use category learning to examine what features of intonation contours are particularly important for children and adults speaking different languages. This could inform us about the biases they may bring to second language learning. Second, given the working memory challenges posed by temporal integration, students of category learning may want to consider examining categorization of temporally extended patterns of various kinds (e.g., visual motion patterns and melodies as well as prosodic patterns more generally). Biases in learning categories of temporally extended patterns, due to either cognitive style or working memory limitations, may also be explored as a possible explanation for atypical use of such patterns in non-neurotypical populations, such as individuals with autism or amusia. Finally, future work should explore reasons for why adults are more narrow-minded than children in acquiring (some) novel categories, including the roles of cognitive maturation and prior experience.

Acknowledgements

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

References

Aslin, R. N., & Smith, L. B. (1988). Perceptual development. *Annual Review of Psychology*, *39*(1), 435-473.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*, 94-117.

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*, 255-278.

Bates, D., & Maeschler, M. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-35. Retrieved from http://CRAN.R-project.org/package=lme4.

Berger, C., & Hatwell, Y. (1996). Developmental trends in haptic and visual free classifications: Influence of stimulus structure and exploration on decisional processes. *Journal of Experimental Child Psychology*, *63*(3), 447-465.

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*, 592–597.

Blair, M., & Homa, D. (2003). As easy to memorize as they are to classify: The 5–4 categories and the category advantage. *Memory & Cognition*, *31*(8), 1293-1301.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341-345.

Boyd, J. K., & Goldberg, A. E. (2012). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*, *39*(03), 457-481.

Buss, E., He, S., Grose, J. H., & Hall III, J. W. (2013). The monaural temporal window based on masking period pattern data in school-aged children and adults. *The Journal of the Acoustical Society of America, 133*(3), 1586-1597.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481-495.

Carvalho, P. F., & Goldstone, R. L. (2012). Category structure modulates interleaving and blocking advantage in inductive category acquisition. *Proceedings of the Annual Conference of the Cognitive Science Society, 34*, 186-191.

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*, 165–182.

Choudhury, A., & Kaiser, E. (2012). Prosodic focus in Bangla: A psycholinguistic investigation of production and perception. *Linguistic Society of America Meeting Extended Abstracts 2012*, http://elanguage.net/journals/lsameeting/issue/view/335

Clark, E. V., & Clark, H. H. (1978). Universals, relativity, and language processing. In: J. H. Greenberg (Ed.), *Universals of human language*, Vol. I. (pp. 225–277). Stanford: Stanford University Press.

Cohen, A., & ‘t Hart, J. (1967). On the anatomy of intonation. *Lingua*, *19,* 177-92.

Cristia, A. & Peperkamp, S. (2012). Generalizing without encoding specifics. Infants infer phonotactic patterns on sound classes. In: A. Biller, E. Chung & A. Kimball (eds.) *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp.126-38). Somerville, MA: Cascadilla Press.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vrecken, O. (1996). The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. In: H. T. Bunnell & W. Idsardi (Eds.), *Proceeding of Fourth International Conference on Spoken Language Processing ICSLP 96* (pp. 1393-1396). IEEE.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America*, 66, 12-21.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*, 245-264.

Fox, A. M., Reid, C. L., Anderson, M., Richardson, C., & Bishop, D. V. (2012). Maturation of rapid auditory temporal processing and subsequent nonword repetition performance in children. *Developmental Science, 15*(2), 204-211.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268-294.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32-41.

Glezer, L. S., Jiang, X., & Reisenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. *Neuron*, 62, 199–204.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251-279.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*(1), 585-612.

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, *24*(2), 87-92.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Gussenhoven, C., Repp, B., Rietveld, T., Rump, W., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America, 102*, 3009-3022.

Hart, J. 't, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Hartley, D. E., Wright, B. A., Hogan, S. C., & Moore, D. R. (2000). Age-related improvements in auditory backward and simultaneous masking in 6-to 10-year-old children. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1402-1415.

Hayes, B. (1995). *Metrical stress theory*. Chicago, IL: University of Chicago Press.

Hill, P. R., Hartley, D. E., Glasberg, B. R., Moore, B. C., & Moore, D. R. (2004). Auditory processing efficiency and temporal resolution in children and adults. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1022-1029.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception & Performance*, *37*, 1939-56.

Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six-to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, *41*(1), 84-110.

Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3- to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science, 39,* 268-306.

Kapatsinski, V. (2014). What is grammar like? A usage-based, constructionist perspective. *Linguistic Issues in Language Technology, 11,* 1-41.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the Annual Conference of the Cognitive Science Society*, *34,* 647-652.

Labov, W. (2010). *Principles of linguistic change. Vol. III. Cognitive and cultural factors*. Malden, MA: Wiley-Blackwell.

Lawson, C. A. (2014) When diverse evidence is (and isn't) inductively privileged: The influence of evidence presentation on children's and adults' generalization. *Proceedings of CogSci*, 2537-2542.

Lawson, C. A., & Fisher, A. V. (2011). It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization. *Journal of Experimental Child Psychology, 110*(4), 499-519.

Lewandowsky, S., Yang, L. X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 881-904.

Liberman, M. (1978). The intonational system of English. Ph.D. Thesis, MIT.

Linzen, T., & Gallagher, G. (2014). The timecourse of generalization in phonotactic learning. *Proceedings of the Annual Meetings on Phonology*, *1,* 1-12.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*, 1357-1372.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19-40.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 128-148.

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H. G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, *28*(4), 1042-56.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1,* 3-36.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*, 122-34.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets?. *Memory & Cognition*, *6*(4), 462-472.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*(6), 1113-1126.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: a comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1518-1533.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775-99.

Mitterer, H., Chen, Y., & Zhou, X. L. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science, 35*, 184-197.

Moore, R., Cowan, J. A., Riley, A., Edmonson-Jones, A. M., & Ferguson, M. A. (2011). Development of auditory processing in 6- to 11-yr-old children. *Ear and Hearing*, *32,* 269-285.

Moore, J. K., & Guan, Y. L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology*, *2*(4), 297-311.

Moore, J. K., & Linthicum Jr, F. H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology*, *46*(9), 460-478.

Narasimhan, B., & Dimroth, C. (2008). Word order and information status in child language. *Cognition*, *107*, 317-29.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–38.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*, 39-57.

Nosofsky, R. M. (2000). Exemplar representation without generalization? Comment on Smith and Minda's (2000) "Thirty categorization results in search of a model". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1735-1743.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60,* 355-376.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017-1045.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167-193.

Odom, R. D., & Cook, G. L. (1984). Perceptual sensitivity, integral perception, and similarity classifications of preschool children and adults. *Developmental Psychology, 20,* 560-567.

Oh, E. L., Wightman, F., & Lufti, R. A. (2001). Children's detection of pure-tone signals with random multitone maskers. *Journal of the Acoustical Society of America, 109*, 2888-2895.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(24), 8776-8780.

Pauen, S. (2002). The global-to-basic shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioural Development*, *26*, 492-499.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 373-79.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*(Doctoral dissertation, Massachusetts Institute of Technology).

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce* (pp. 11-36). Dordrecht: Kluwer.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 143-170.Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77,* 353-363.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org

Rabi, R., & Minda, J. P. (2014a). Rule-based category learning in children: The role of age and executive functioning. *PloS One*, *9*(1), e85316.

Rabi, R. R., & Minda, J. P. (2014b). Perceptual category learning: Similarity and differences between children and adults. *Proceedings of CogSci 36.*

Reed, S. K. (1978). Category vs. item learning: Implications for categorization models. *Memory & Cognition*, *6*(6), 612-621.

Reinisch, E., & Weber, A. (2012). Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America*, *132*(2), 1165-1176.

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories?. *Journal of Phonetics, 45,* 91-105.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A Parallel Distributed Processing approach*. Cambridge, MA: MIT Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, *125*, 49–63.

Sag, I., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. *Chicago Linguistic Society, 11,* 487-497.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science*, *6*(5), 557-567.

Schneider, B. A., Trehub, S. E., Morrongiello, B. A. & Thorpe, L. A. (1989). Developmental changes in masked thresholds. *The Journal of the Acoustical Society of America, 86*, 1733-1742.

Schwarzer, G. (1997). Analytic and holistic modes in the development of melody perception. *Psychology of Music*, *25*(1), 35-56.

Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication, 66*, 65-81.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, *25*(4), 425-437.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 1-42.

Shport, I. A. (2011). *Cross-linguistic perception and learning of Japanese lexical prosody by English listeners* (Doctoral dissertation, University of Oregon).

Siegenthaler, B. M. (1969). Maturation of auditory abilities in children. *International Journal of Audiology, 8*, 59-71.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3-27.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411-36.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*(1), 125-144.

Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science*, *22*(8), 1049-1057.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, *1*(1), 85-123.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Science, 43*, 454-64.

Tahta, S., Wood, M., & Loewenthal, K. (1981). Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech*, *24*(4), 363-372.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development, 65*, 1627-1645.

Thompson, L. A., & Massaro, D. W. (1989). Before you see it, you see its parts: Evidence for feature encoding and integration in preschool children and adults. *Cognitive Psychology*, *21*(3), 334-362.

Tremblay, A. (2009). *Processing advantages of lexical bundles: Evidence from self-paced reading, word and sentence recall, and free recall with event-related brain potential recordings* (Doctoral dissertation, University of Alberta).

Walsh, M., Schweitzer, K., & Schauffler, N. (2013). Exemplar-based pitch accent categorisation using the Generalized Context Model. In *INTERSPEECH* (pp. 258-262).

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family resemblance concepts. *Memory & Cognition, 15,* 45-52.

Ward, T. B., Vela, E., & Hass, S. D. (1990). Children and Adults Learn Family-Resemblance Categories Analytically. *Child Development*, *61*(3), 593-605.

Weinberger, S. (2013). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu

Werner, L. A., & Boike, K. (2001). Infants' sensitivity to broadband noise. *The Journal of the Acoustical Society of America*, *109*(5), 2103-2111.

Wills, A. J., Milton, F., Longmore, C. A., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification?. *The Quarterly Journal of Experimental Psychology*, *66*(2), 299-318.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*, 209–212.

Wright, M., Green, A., & Baker, S. (2000). Limitations for change detection in multiple Gabor targets. *Visual Cognition*, *7*, 237-252.

Wundt, W. M. (1900). *Die Sprache*. Leipzig: Engelmann.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245-272.

Figure 1. The surprise redundancy contour used to indicate surprising information for *The blackboard is orange*. and *The canoe is orange*, produced by a native speaker of American English. H(igh) and L(ow) are tone values. Starred tones are aligned with stressed syllables (*black* in *blackboard*, *noe* in *canoe,* and *o* in *orange*). If the same utterances were pronounced 'normally', as simple declarative statements, the initial HL* sequence would have been replaced by a single tone of intermediate height aligned with the beginning of the utterance, resulting in a gradual rise to the H* on the 'o' of *orange*.

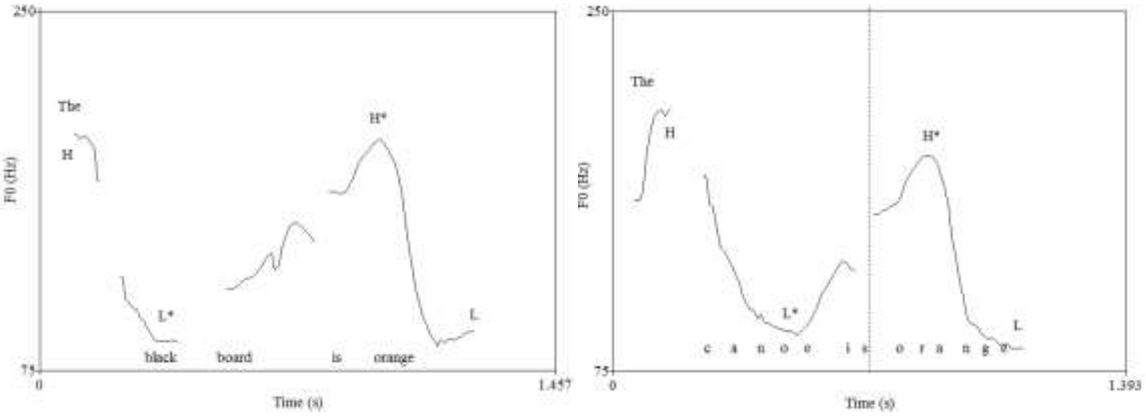

Figure 2. Pitch contours used in testing. Black dashed line = prototypes. Magenta = one-semitone distortions (four of which were used in training) and, for M, the distractors. Blue = three-semitone distortions. Green = five-semitone distortions.

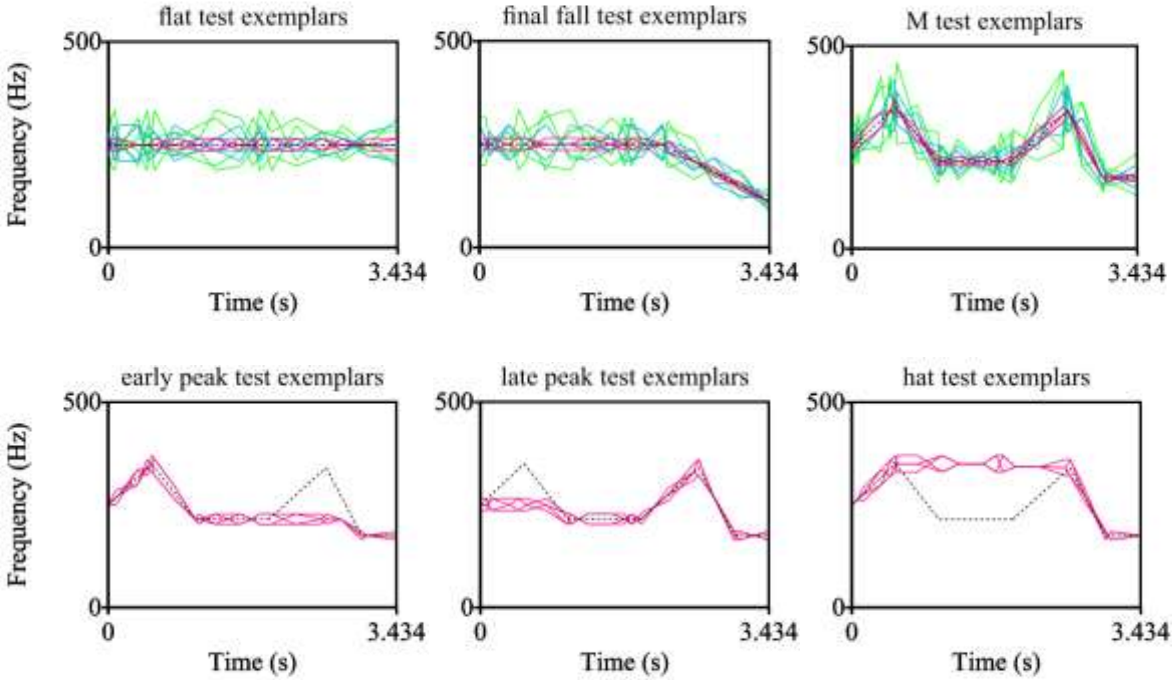

Figure 3. The test response screen.

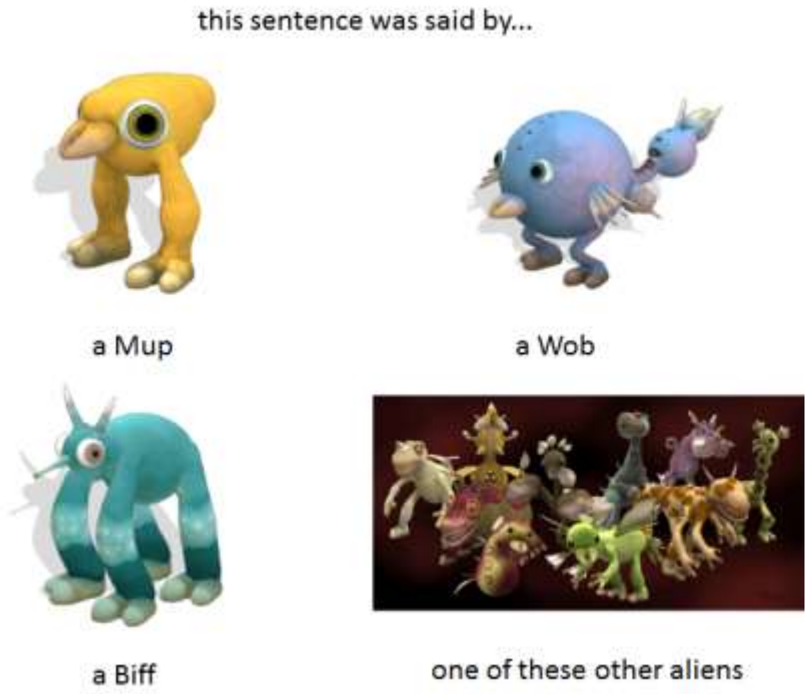

Figure 4. Acceptance of deviations by distortion level and age group. (TR1 = previously experienced exemplars, ST = semitones by which the exemplars can maximally deviate from the prototype). Error bars are 95% confidence intervals.

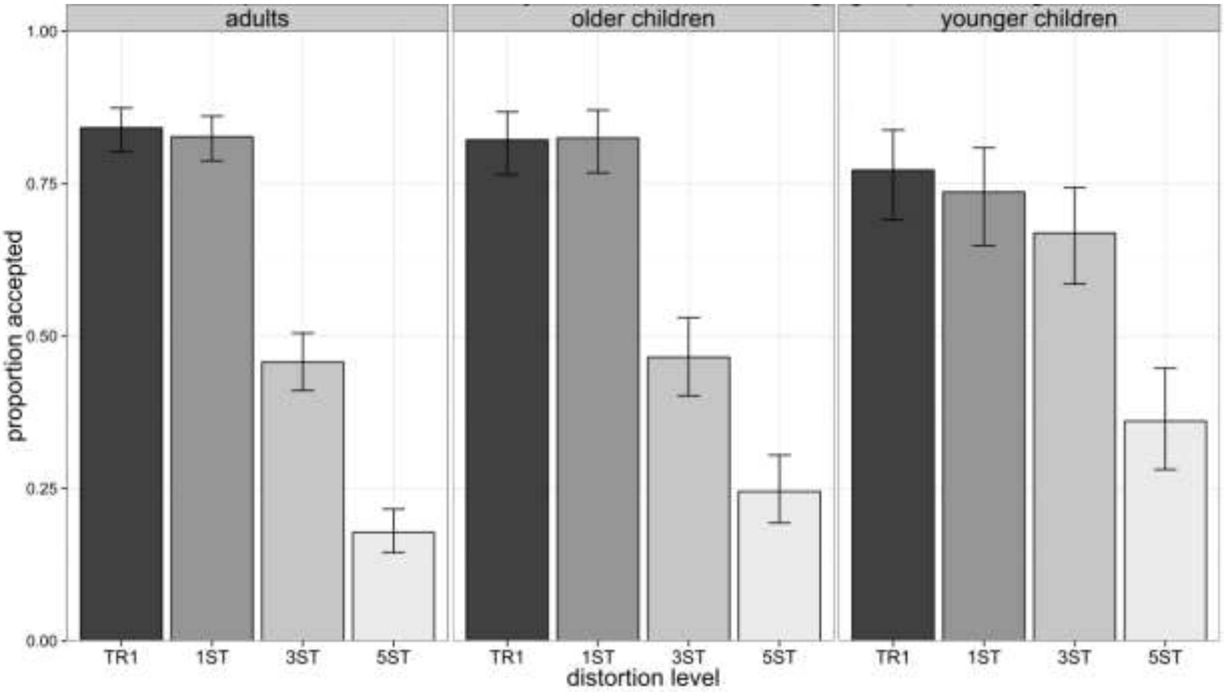

Figure 5. Acceptance of early peak (/____), late peak (___/_), and hat (/¯¯¯_) distractors and new low-level (1 semitone) distortions of M (/_/\\ _) by age group. Error bars are 95% confidence intervals based on the proportion test.

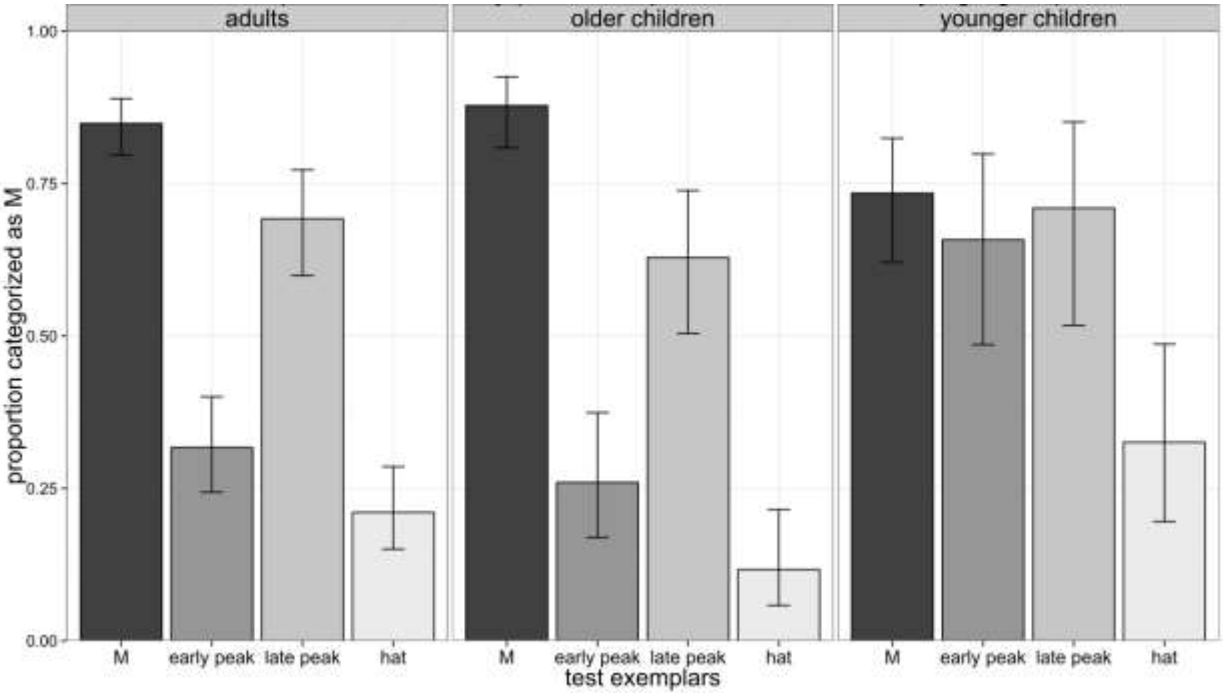

Table 1. The interaction between group (child vs. adult) and distance from the prototype (where novel 1 semitone distortions are the reference level) on acceptability (1 = 'accepted', 0 = 'rejected') for novel distortions. Negative coefficients = lower acceptability, positive coefficients = higher acceptability.[9]

	Coefficient	**Std.Error**	***z***	***p***
Distance:				
3 semitones	-2.0585	0.2934	-7.015	<.0001***
5 semitones	-3.7993	0.3390	-11.206	<.0001***
Group: child	-0.2658	0.3078	-0.864	.39
Distance x Group:				
3(st):child	0.7027	0.3206	2.192	.028*
5(st):child	1.1773	0.3937	2.990	.0028**

---

[9] The reported model, including the group by distortion level interaction, is preferred to a simpler model that does not include the interaction according to the log likelihood test ($\chi^2(2)=8.17$, p = .017)

Table 2. The interaction between continuous age and distance from the prototype (where novel 1 semitone distortions are the reference level) on acceptability (1 = ‘accepted’, 0 = ‘rejected’) for children presented with novel distortions. Age was standardized. Negative coefficients = lower acceptability, positive coefficients = higher acceptability.[10]

	Coefficient	**Std.Error**	***z***	***p***
Distance:				
3 semitones	-1.2622	0.2822	-4.473	<.0001***
5 semitones	-2.5568	0.3031	-8.436	<.0001***
Age	0.1618	0.2053	0.788	.43
Distance x Age:				
3(st):Age	-0.6584	0.2334	-2.821	.005**
5(st):Age	-0.5584	0.2602	-2.146	.03*

---

[10] The reported model, including the age by distortion level interaction, is preferred to a simpler model that does not include the interaction according to the log likelihood test ($\chi^2(2)=7.16$, $p = .028$).

Table 3. The interaction between age (median split) and distance from the prototype: Only younger children significantly differ from adults in the effect of distance on acceptance of novel distortions.[11]

	Coefficient	**Std.Error**	***z***	***p***
Distance:				
3 semitones	-2.06126	0.28781	-7.162	<.0001***
5 semitones	-3.79664	0.33398	-11.368	<.0001***
Age (categorical):				
>10;3 (Older)	-0.12284	0.37569	-0.327	.74
<10;3 (Younger)	-0.4408	0.38401	-1.148	.25
Distance x Age (categorical):				
3(st):Older	0.08406	0.37508	0.224	.82
5(st):Older	0.63351	0.46518	1.362	.17
3(st):Younger	1.41898	0.39428	3.599	.0003***
5(st):Younger	1.74912	0.47743	3.664	.0002***

---

[11] The reported model, including the age category by distortion level interaction, is preferred to a simpler model that does not include the interaction according to the log likelihood test ($\chi^2(4)=15.96$, p = .003). At the same time, the interaction between age category and distortion level is not siginificant if the younger children are excluded, in agreement with the lack of significant for 3(st):older and 5(st):older in Table 3 ($\chi^2(2)=3.05$, p = .22).

Table 4. The effect of subject group on acceptance of distractors lacking one of the features of M (/_/_) as being instances of M. Novel 1-semitone distortions of the M prototype are the reference level to which the distractors are compared. Positive coefficients indicate greater acceptability. Negative coefficients indicate lower acceptability.[12]

	Coefficient	**Std.Error**	***z***	***p***
Contour:				
/____ (early peak)	-4.809	0.7033	-6.837	<.0001***
___/_ (late peak)	-2.7566	0.6267	-4.399	<.0001***
/¯¯¯¯_ (hat)	-5.576	0.7513	-7.422	<.0001***
Group: child	-1.6404	0.8955	-1.832	0.067
Contour x Group:				
/____:Child	2.1065	0.908	2.32	.02*
___/_:Child	1.3546	0.7966	1.701	.09
/¯¯¯¯_:Child	1.4082	0.9954	1.415	.16

---

[12] The interaction between distractor type and group does not come out significant in an omnibus test ($\chi^2(3)=5.96$, $p = .11$), since the only distractor type that children as a group seem to prefer compared to adults are early peak distractors.

Table 5. The effect of continuous age for the 9- to 11-year-old children on acceptance of distractors lacking one of the features of M (/_/_) as being instances of M. Novel 1-semitone distortions of the M prototype are the reference level to which the distractors are compared. Positive coefficients indicate greater acceptability. Negative coefficients indicate lower acceptability.[13]

	Coefficient	**Std.Error**	***z***	***p***
Contour:				
/____ (early peak)	-2.3191	0.4862	-4.77	<.0001***
___/_ (late peak)	-1.0985	0.5149	-2.134	.03*
/¯¯¯¯_ (hat)	-3.886	0.5761	-6.746	<.0001***
Age	0.6398	0.4436	1.442	.15
Contour x Age:				
/____:Age	-1.7562	0.4764	-3.687	.0002***
___/_:Age	-1.032	0.4962	-2.08	.04*
/¯¯¯¯_:Age	-1.3015	0.5556	-2.342	.02*

---

[13] The interaction between distractor type and group is significant in an omnibus test (χ^2(3)=11.22, *p* = .01).

Table 6. The effect of subject group on acceptance of distractors lacking one of the features of M (/_/_) as being instances of M. Novel 1-semitone distortions of the M prototype are the reference level to which the distractors are compared. Positive coefficients indicate greater acceptability. Negative coefficients indicate lower acceptability.[14]

	Coefficient	**Std.Error**	***z***	***p***
Contour:				
/____ (early peak)	-4.24798	0.57325	-7.41	<.0001***
___/_ (late peak)	-2.32139	0.51335	-4.522	<.0001***
/¯¯¯¯_ (hat)	-5.06802	0.63184	-8.021	<.0001***
Group: Older child	-0.76274	0.97195	-0.785	.43
Group: Younger child	-1.88037	0.9035	-2.081	.04*
Contour x Group:				
/____:Older	-0.01344	0.9745	-0.014	.99
___/_:Older	0.31018	0.86428	0.359	.72
/¯¯¯¯_:Older	-0.49484	1.13352	-0.437	.66
/____:Younger	3.59959	0.89359	4.028	<.0001***
___/_:Younger	1.80596	0.7869	2.295	.02*
/¯¯¯¯_:Younger	2.56488	0.99887	2.568	.01*

---

[14] The age group / distractor type interaction is likewise significant: ($\chi^2(6)=15.34$, $p = .018$).

Appendix: Stimulus type / Response cross-tabulations. Stimulus categories are along the top, while responses are labeled in the leftmost column. Most frequent responses for each stimulus type for any given participant group are bolded.

Table A1. Familiar exemplars for adults

	final fall	flat	M
trash	23	31	12
final fall	**131**	2	7
flat	2	**104**	36
M	4	23	**105**

Table A2. Novel low-level (1 st) distortions for adults, including low-level distortions of M-distractor prototypes

	final fall	flat	M	early peak	late peak	hat
trash	18	41	16	**95**	36	**112**
final fall	**134**	4	8	4	8	7
flat	6	**102**	38	19	43	9
M	2	13	**98**	42	**73**	32

Table A3. Novel mid-level (3 st) distortions for adults

	final fall	flat	M
trash	**78**	**82**	**71**
final fall	73	1	3
flat	7	57	23
M	2	20	63

Table A4. Novel high-level (5 st) distortions for adults

	final fall	flat	M
trash	**115**	**125**	**129**
final fall	30	0	2
flat	8	25	6
M	7	10	23

Table A5. Familiar exemplars for older children

	final fall	flat	M
trash	10	22	6
final fall	**58**	1	4
flat	4	**47**	14
M	8	10	**56**

Table A6. New low-level distortions for older children, including low-level distortions of M-distractor prototypes

	final fall	flat	M	early peak	late peak	hat
trash	8	23	7	**54**	22	**61**
final fall	**64**	2	7	3	2	7
flat	1	**44**	17	10	17	6
M	7	11	**49**	13	**39**	6

Table A7. New mid-level distortions (3 st) for older children

	final fall	flat	M
trash	36	**43**	**38**
final fall	**40**	0	3
flat	2	26	9
M	2	11	30

Table A8. New high-level distortions (5 st) for older children

	final fall	flat	M
trash	**54**	**55**	**56**
final fall	20	2	3
flat	1	14	4
M	5	9	17

Table A9. Familiar exemplars for younger children

	final fall	flat	M
trash	12	17	5

final fall	**51**	9	13
flat	9	**34**	14
M	8	20	**48**

Table A10. New low-level distortions for younger children, including low-level distortions of M-distractor prototypes

	final fall	flat	M	early peak	late peak	hat
trash	8	16	11	16	13	**36**
final fall	**52**	11	13	13	17	11
flat	11	**36**	21	19	23	16
M	9	17	**35**	**32**	**27**	17

Table A11. New mid-level distortions (3 st) for younger children

	final fall	flat	M
trash	16	25	21
final fall	**46**	8	10
flat	11	**37**	17
M	7	10	**32**

Table A12. New high-level distortions (5 st) for younger children

	final fall	flat	M
trash	25	**44**	**38**
final fall	**28**	11	9
flat	14	14	17
M	13	11	16