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 that are expected to pose significant working memory challenges if represented in their full detail. We report that both children and adults form relatively abstract representations of intonation contours: previously encountered and novel exemplars are categorized into a category equally often, as long as distance from the prototype is controlled. However, we also document age-related differences in categorization performance. Given the same experience, adults form narrower categories than children, which may make it harder for them to recognize new instances of a contour as belonging to a known category. In particular, children are more likely to accept examples that lack some of the characteristic features of the prototype as belonging to the category. Children are also more likely to accept distortions of the prototype that are less similar to the prototype than the training examples. There are also age-related differences in feature weight: while adults pay more attention to the end of the contour, children pay approximately equal attention to the beginning and the end of the contour. 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. We discuss implications for theories of categorization, including their applicability to categories of linguistic units, and challenges posed by intonational patterns to various learner populations.

Keywords: prosody, intonation, categorization, perceptual learning, inductive bias, starting small, second language acquisition, sensitive periods

1. Introduction

The present paper explores perceptual learning of intonation contour categories in children and adults. Perceptual learning in speech refers to the formation of categories of acoustically and/or visually different stimuli, such as the different pronunciations of a speech sound (e.g. Maye et al. 2002, Norris et al. 2003, Bertelson et al. 2003). Perceptual learning is a type of category learning (e.g. Posner & Keele 1968) that is distinguished by the fact that the learned categories are thought to be useful for future perceptual processing (e.g. Goldstone 1998, p.585, defines it as “relatively long-lasting changes to an organism’s perceptual system that improve its ability to respond to its environment”).

So far, work on the perceptual learning of language has focused 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. 2013, 2014, van der Zande et al. 2013). Recent work has started extending perceptual learning to local suprasegmental patterns: lexical tone distinctions instantiated over single syllables (Francis et al. 2008, Maddox et al. 2013, Maddox & Chandrasekaran 2014, Mitterer et al. 2011), as well as pitch accents (Shport 2011), and stress patterns (Reinisch & Weber 2012) instantiated over pairs of adjacent syllables. Work on non-linguistic category learning has likewise focused on stimuli with little to no temporal extent (static images). We are aware of only two studies that have studied how children and adults acquire categories of temporally extended non-linguistic patterns (Berger & Hatwell, 1996, on haptically explored objects, and Schwarzer, 1997, on melodies). The present study thus extends work on perceptual learning of language to the acquisition of temporally extended suprasegmental patterns (see also Kurumada et al. 2012) whose perception and integration may pose challenges to working memory similar to those documented for syntactic constructions (e.g. Bartek et al. 2011, Fiebach et al. 2001, Hofmeister & Sag 2010, Kluender & Kutas 1993, Lewis et al. 2006, Sagarra & Herschensohn 2010). In this introductory section, we aim to convince the reader that perceptual learning of intonation contours is an interesting research problem for students of categorization, working memory, first and second language acquisition and atypical speech/language development.

What are intonation contours?

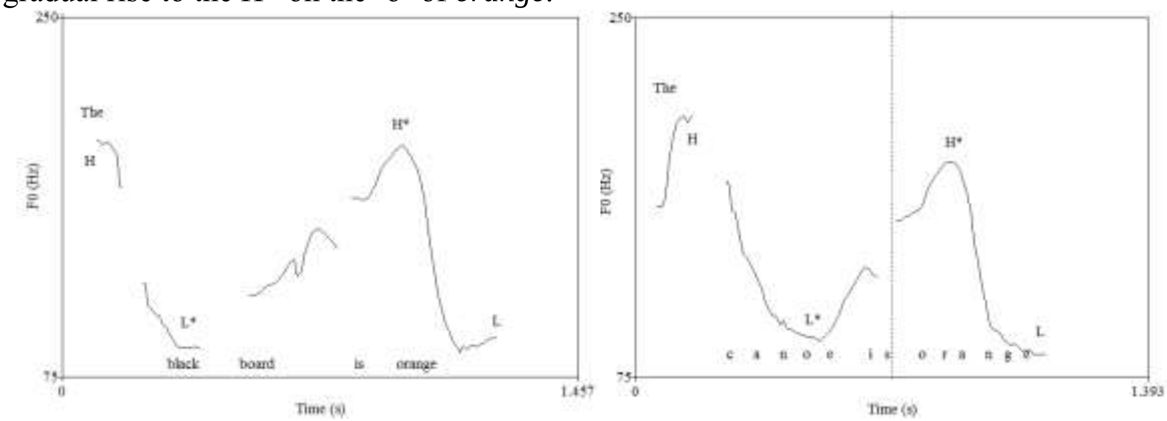
Intonation contours are meaningful pitch patterns distributed over utterances (e.g. ‘t Hart et al. 1990). Changing the intonation contour of an utterance can make the difference between a statement and a question. It can also cue that some of the information in that utterance is surprising or predictable. An example of an intonation contour can be seen in Figure 1. The most unambiguous evidence for categories of entire contours comes from the idiomatic value of some contours such as the so-called surprise redundancy contour (Figure 1, first described in Sag & Liberman 1975): a contour can mean more than the sum of its parts (see also Gorisch et al., 2012; Hedberg et al., 2006, N. Ward 2014).

The dimensions[1] that define the space of intonation contours are commonly believed to include relative pitch values (Low or High) of the contour's inflection points (Bruce & Gårding 1978, Hayes 1995, Ladd 1983, Liberman 1978, Pierrehumbert 1980, 2000, Pike 1945), which are referred to as tones (L or H), the slopes of upglides and downglides between the inflection points ('t Hart et al. 1990) or both (Cho 2010). Accordingly, the category can be defined as a sequence of tonal targets (inflection points) and/or as a sequence of pitch movements (slopes). In the tonal approach, the surprise redundancy contour is often described as the tonal sequence H L* H* L (e.g. Hayes 1995: 16-18), where * represents tones aligned to stressed syllables in critical words and % represents tones aligned with a phrase boundary. The tonal targets represent inflection points in the contour where the pitch movement trajectory changes direction. As one can see in Figure 1, the temporal intervals between these tonal features are variable and can be quite long (see also ‘t Hart et al. 1990, pp.82-88).

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

[1] The terms *dimension*, *feature* and *property* are used somewhat interchangeably, except dimensions are typically continuous whereas features are typically discrete.

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*.[2]



There is disagreement in the literature on whether intonation contours are morphemes, undecomposable into smaller meaningful elements (L. Armstrong & I. Ward 1926; Gorisch et al., 2012; Hedberg et al., 2006; Jones, 1957; Liberman & Sag, 1974; Sag & Liberman 1975, N. Ward 2014; Wennerstrom 2001), or sequences of simpler tonal morphemes (Gussenhoven 2004, Pierrehumbert & Hirschberg 1990). This may lead some to wonder whether intonation contours are plausibly thought of as perceptual categories. However, research on the processing of morphology suggests that extended forms that are treated by analysts as decomposable are often processed by listeners as single units. 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). For example, clearly compositional multimorphemic words and even phrases show signs of whole-unit storage and retrieval as long as they are frequently used (e.g. Alegre & Gordon 1999, Baayen et al. 1997, Kapatsinski & Radicke 2009, Sereno & Jongman 1997, Tremblay 2009). In particular, Tremblay (2009) shows that frequency of a 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 hypothesizes *in the middle of* as early as the beginning of *in* and would therefore profit from learning the acoustic characteristics of *in the middle of*. These acoustic characteristics would include both early and late cues to its occurrence, allowing for robust recognition in noise (e.g. Salasoo & Pisoni 1985). Summarizing decades of work from a bottom-up perspective on intonation in which contours are distinguished based on perceptual rather than semantic equivalence, 't Hart et al. (1990) write that "it is the intonation pattern that dictates the local choice among the various pitch movements" (p.66), and define intonation patterns as

[2] See, e.g. Fujisaki (1983), Mishra (2008), van Santen et al. (2005), Thorsen (1978) for alternatives based on decomposing contours into superimposed contour curves of different temporal spans (e.g. phrases and feet). Despite the promise of these approaches for speech synthesis, given the largely incremental nature of perceptual processing, relatively local features still appear needed to serve as cues to the presence of the various contours, and it is clear from empirical work on intonation perception that some parts of contours are perceptually more important than others (e.g. Gussenhoven et al. 1997, 't Hart et al. 1990), just as some parts of images draw more attention than others (e.g. Parkhurst et al. 2002).

"melodic categories… characterized by a unique combination of invariant abstract melodic properties… instantiated in a great variety of pitch contours" (pp.83-84).[3]

Thus, we maintain that intonation contours that share a meaning (e.g. that the utterance is a question) can be seen as a category, and in the present study examine intonation contour learning as formation of a perceptual category along the lines of Posner & Keele (1968). All known languages have intonation (Bolinger 1964). However, specific intonational patterns vary across languages 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 (Grabe et al. 2003). In the rest of this introduction we discuss what needs to be learned to form an intonation contour category from perceptual experience, the inductive biases that may be expected to shape the process, and possible developmental changes in these biases.

Relevance to models of categorization

Intonation contour categories exhibit structures that have been underexplored in the literature on category learning (e.g. Murphy 2002, J. Smith & Minda 2000) and have the potential to shed light on central questions in the field, including the ubiquity of reliance on memorized exemplars of previously encountered stimuli across category structures (e.g. Minda & J. Smith 2001, Rouder & Ratcliff 2004), and the relationship between categorization and working memory (e.g. Lewandowsky et al. 2012, Minda et al. 2008). As reviewed in J. Smith & Minda (2000) and Minda & J. Smith (2001), most studies have examined small categories of low-dimensional visual stimuli, in which the individual members of the categories can be easily memorized (Blair & Homa 2003, Reed 1978, Rouder & Ratcliff 2004). Furthermore, the participants in most studies have been trained to criterion on categorizing the training stimuli, again encouraging memorization (J. Smith et al. 1997). The categories in most studies have also been difficult to separate while the exemplars within the categories are easily distinguishable from each other, which makes it difficult to achieve accurate categorization without memorizing specific exemplars (Minda & J. Smith 2001). All of these factors have been shown to encourage reliance on memorizing specific exemplars of category members presented during training (Homa et al. 1979, 1981, Lewandowsky et al. 2012, Minda & J. Smith 2001, Reed 1978, Rouder & Ratcliff 2004, J. Smith & Minda 1998, J. Smith et al. 1997).

In contrast, there is a huge variety in possible realizations of an intonation contour (e.g. 't Hart et al. 1990). Every time a contour is pronounced, it is pronounced slightly differently; in part due to peculiarities of individual speakers, in part due to segmental differences in the utterances over which the contours are superimposed. Thus, a category of contours sharing meaning would include a huge number of individual contour exemplars. Contour exemplars are highly multidimensional, involving a long pitch movement trajectory, and thus should be relatively difficult to memorize. At the same time, intonation contour categories are characterized by multiple redundant features, which have variable realizations but are almost always realized in instances of a contour (e.g. Liberman 1978, Pierrehumbert 1980, 2000, 't Hart et al. 1990; but cf. Watson et al. 2008). For example, recall that the surprise-redundancy contour seen in Fig.1 is characterized by the sequence HL*H*L%, but the individual pitch values and spacing between the tonal targets is variable. As noted in Kapatsinski (2014), this kind of

[3] Of course, viewing instances of a contour as a category does not preclude the possibility that this category is defined by a single feature/property, rather than a combination of features. It simply makes this an empirical question.

structure is typical of meaningful categories of linguistic forms (*constructions*). For example, if I spell the word *monotonous* as *conotonous*, it would no longer be perceived as *monotonous* despite sharing almost all features with it, unless the listener perceptually undoes the change, thinking that the *c* is just an erroneous realization of an intended *m*. The existence of redundant features that characterize all or most instances of a contour may encourage reliance on individual features rather than memory for whole exemplars. T. Ward et al. (1990) point out that a category characterized by several redundant features allows learners with different cognitive styles or prior experiences to focus on different features and yet converge on the same categorization. This behavioral uniformity in the face of perceptual and cognitive diversity is particularly important for categories of linguistic forms, which need to be stable at the behavioral level across the speech community for reliable communication to be possible.[4]

Acquisition of intonation: A perceptual learning challenge?

Acquisition of intonation poses a challenging problem for second language learners, particularly when a native speaker of a language with lexical tones, like Mandarin, tries to acquire a language without lexical tones, like English, or vice versa (e.g. Anderson-Hsieh et al. 1992, Bent 2005, Grabe et al. 2003, Makarova 2001, Shport 2011, Yang & Chan 2010). Acquisition of intonation is also challenging for first language learners with autism spectrum disorders (e.g. Diehl et al. 2008, McCann et al. 2007, Paul et al. 2005, Peppé et al. 2007, Shriberg et al. 2001).

We hypothesize that in order to perceive an intonation *contour*, one needs to integrate the tonal features defining the contour. The ability to access several "information elements" simultaneously and integrate them into larger structures has recently been argued to be an important component of working memory (Halford et al. 1998, Oberauer et al. 2000, 2003, Robin & Holyoak 1995). Perceiving an intonation contour thus requires *temporal integration* in working memory: the tonal features of a contour that are to be integrated are presented sequentially. This contrasts with previous studies of the relationship between working memory and category learning in which the features of a stimulus were presented simultaneously, thus requiring little or no temporal integration (Lewandowsky et al. 2012, Minda et al. 2008, Rabi & Minda 2014).

Temporal integration is an ability that has been observed to vary amongst individuals (for tones, see Haywood et al. 2011, Vinnik et al. 2012). While individual differences in tone integration are not yet explained (Haywood et al. 2011), Vinnik et al. (2012) suggest that decreased integration may be attributable to increased attention to local acoustic detail (which they find to vary within and across individuals). Increased attention to detail and reduced ability to integrate the details into a perceptual whole have been suggested to characterize perception and cognition in individuals with autism (Frith 1989, Frith & Happe 1994). The expected difference between neurotypical and autistic participants in temporal integration was recently demonstrated by Nakano et al. (2010) for visual features. It was subsequently replicated by Falter et al. (2012), who also found a correlation between differences in visual temporal integration and communication skill, suggesting that the temporal integration deficits might be domain-general. Based on these findings, children with autism may be expected to face difficulties in acquiring

[4] A visual analogy may be a dot moving up and down on the screen, where there is an invariant sequence of ups and downs but the magnitudes and timings of these movements are variable. Of course, such stimuli might pose even greater challenges to working memory, as keeping track of visual sequences appear to be harder than keeping track of auditory ones (Conway et al. 2009).

the intonation contours of their native language, insofar as perception of an intonation contour requires temporal integration.

As previously noted, intonation contours are long trajectories of pitch movement, often modeled as sequences of alternating values (High and Low) of a single feature (e.g. Liberman 1978, Pierrehumbert 1980, 2000). Working memory capacity has been found to be reduced when the stimuli to be kept in mind and integrated share features (Nairne 1990, Oberauer & Kliegl 2006, Saito & Miyake 2004). Being different values of the same feature, the to-be-integrated tones of an intonation contour thus share all features and should be difficult to integrate. Acquisition of intonation contour categories is therefore expected to be especially challenging for working memory.

Learning a *category* of intonation contours that share a meaning also requires the ability to deal with within-category variability (a major problem in speech perception generally, e.g. Pisoni & Levi 2007). Competing influences on pitch -- including emotional affect, lexical stress, and, in languages like Mandarin, lexical tone -- can make instances of the same contour sound very different from each other. Yang (2011) and Yang & Chan (2010) find that American learners of Chinese have trouble distinguishing Mandarin questions and statements in the presence of local pitch variability caused by lexical tone. From an English speaker's perspective, the local pitch variability characteristic of Mandarin is unexpected and therefore attention-demanding, hence different instances of the Mandarin question intonation may not be perceptually similar enough to be grouped into a single category. Conversely, Mandarin learners of English may tolerate a much larger range of local pitch variability within instances of the same contour due to experience with broad first language intonation contour categories. Thus English listeners judge local, syllable-to-syllable, pitch variability to be the most salient characteristic of a Mandarin accent (Anderson-Hsieh et al. 1992).

The present study reports on intonation contour learning in monolingual English-speaking adults and older children, in part to provide a baseline to which intonation contour learning in second language learners and individuals with ASD can be compared.

Developmental differences in category breadth: Do children start small or start big?

One of the central challenges for language acquisition is to explain why children are better than adults at learning languages. There are two contradictory explanations that have posited developmental changes in domain-general inductive biases influencing category breadth, both drawing support primarily from the literature on the acquisition of syntax. One, the "less is more" (Newport 1990) or "starting small" (Elman 1993) hypothesis, suggests that children have helpful cognitive limitations, including, crucially, a smaller working memory capacity. These limitations help children focus on narrower temporal spans of speech, which contain a small number of individual, localized features. This allows children to zero in on the important features, separating the wheat from the chaff, albeit at the expense of temporal integration. Given that an intonation contour category for an adult has several necessary tonal features, a child’s contour category may be broader, as the child has focused on only some of the features of the adult category.

The possibility that children focus on a smaller set of features than adults in acquiring a category has also been raised in foundational work on perceptual learning. Gibson & Gibson (1955) report on an experiment in which children between the ages of 6 and 11 years and adults were repeatedly presented with a single scribble and had to judge other scribbles differing from it in various ways on whether or not they were the same scribble. Of particular relevance to the starting small hypothesis, they reported that:

> For adults, the class of undifferentiated items at the outset was small…, and only a few trials were needed before this class was reduced to the critical item alone… At the other extreme, however, the younger children [6-8 years old] "recognized" nearly all of the scribbles on the first trial…, which is to say that the class of undifferentiated items was large… For the older children (between 8½ and 11 years of age) the results were intermediate between these extremes." (Gibson & Gibson 1955, pp.37-38).

A different perspective, the *holistic processing view*, with contrasting developmental predictions, arose from the work on the acquisition of ill-defined categories (Aslin & L. Smith 1988, Brooks 1978, Kemler Nelson 1984, 1988, Shepp et al. 1980, J. Smith & Shapiro 1989, L. Smith 1989, L. Smith & Kemler 1977, 1978, 1984). This work built on Garner's (1974) distinction between separable and integral dimensions. In Garner's framework, separable dimensions are ones that can be attended separately (e.g. shape and hue of a figure),, whereas integral dimensions cannot because they are internally represented as a single undecomposable dimension (e.g. hue and saturation).. According to Shepp et al. (1980, p.119), "dimensional combinations that are perceived by older children and adults as separable are perceived by younger children as integral". In the limit, rather than attending to a subset of dimensions characterizing a stimulus, the child on this view attends to all dimensions equally (L. Smith 1989). Thus, categorization in adults may be based on a small subset of features characterizing stimuli while children's categorization is based on a larger set of features. As summarized by T. Ward et al. (1990, p.593), the holistic processing view suggests that "the young child's preferred mode of processing is a holistic one, and that such a mode is manifested in and well suited to learning the structure of many naturally occurring categories", particularly ones that are well separated from other categories due to having many redundant characteristic properties, as we have argued to be the case for intonation contours (see also Pierrehumbert 2000). If true, this hypothesis would undermine our assumption that children as well as adults need to integrate multiple features together to perceive a contour. Instead, they could be hypothesized to process contours holistically, without decomposing them into individual features and without requiring integration of those features in working memory.

In the language acquisition literature, the holistic-to-analytic developmental shift has been proposed under the name of the "starting big" hypothesis (Arnon 2009, Arnon & Ramscar 2012; see also Tomasello 2003). The starting big hypothesis suggests that children learn from holistic units that are larger and less analyzed than the ones adults learn from, relying on detailed memories of specific exemplars of utterances instantiating syntactic constructions. In language acquisition, the "starting big" and "starting small" hypotheses have been applied to the acquisition of morphosyntax, and evaluated mostly using production data (see the review in Ambridge & Lieven 2011).. A principal aim of the present study is to evaluate their applicability to the perceptual learning of intonation.[5] Like some syntactic constructions, intonation contours

[5] More broadly, we wish to contribute to the debate on the extent to which categories become more and more specific vs. more and more general over the course of development (see also Albright & Hayes 2003, Kapatsinski 2013,

are extended in time. This characteristic makes them prime candidates for testing between the two hypotheses in a new domain.

While previous work supporting the “starting small” and “starting big” hypotheses for language processing examined very young children (2-year-olds), we suspected that the developmental change in category breadth may continue throughout childhood for several reasons. First, working memory continues to develop into adolescence (Luna et al. 2004). Second, developmental differences in categorization have been obtained with older children (e.g. Berger & Hatwell 1996, Gibson & Gibson 1955, Schwarzer 1997, L. Smith 1989). Third, the 8-11 age range has been suggested to be an important transition time for the ability to acquire complex sound patterns. In particular, Tahta et al. (1981) examined the ability of children aged 5-15 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: 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 phonological patterns of the speech of that community. He notes that this “does not imply that the language learning mechanism declines abruptly at that age” (ibid.). Rather, the language learning mechanism is argued to become adultlike by 11. Thus in this study we examined 9-11 year-olds. 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 are adultlike in their learning biases.

As discussed by 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, again resulting in broader categories. 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.[6]

2. Experiment 1: Category learning

2.1. Methods

This experiment applied the schema abstraction paradigm (Posner & Keele, 1968) to the categorization of intonation contours. Listeners were trained to associate contour exemplars with one of three alien languages. The exemplars consisted of low-level distortions of category prototypes which were themselves withheld during training. At test, listeners were presented with trained exemplars, new low-level distortions, prototypes, mid- and high- level distortions,

Linzen & Gallagher 2014 for segmental phonology; Jenkins et al. 2014, Rogers & McClelland 2004, Xu & Tenenbaum 2003 for word meaning; Fennell & Werker 2003, Stager & Werker 1997, Swingley 2007 for word form).

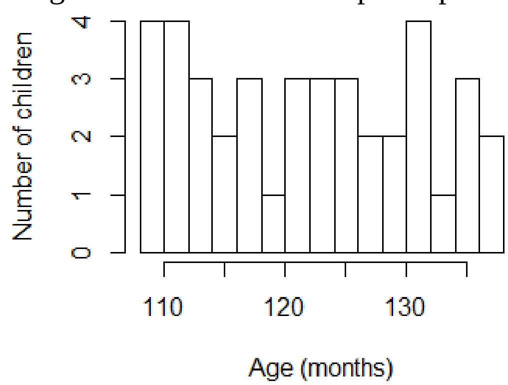
[6] Note that other explanations for age-related differences in category breadth are possible. We return to those in the general discussion.

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 unclassifiable.

Participants

Forty children and forty adults participated in the experiment. The adults were recruited from introductory psychology classes through the INSTITUTION Psychology/Linguistics Human Subject Pool. They received course credit for participating. The children were 9-11 years old and were recruited from local schools, by word of mouth and through a database maintained by the INSTITUTION'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, as shown in Figure 2. The children received a small gift for participating, and families were compensated for their time.

Figure 2. A histogram of the ages of the children who participated in the experiment



Stimuli

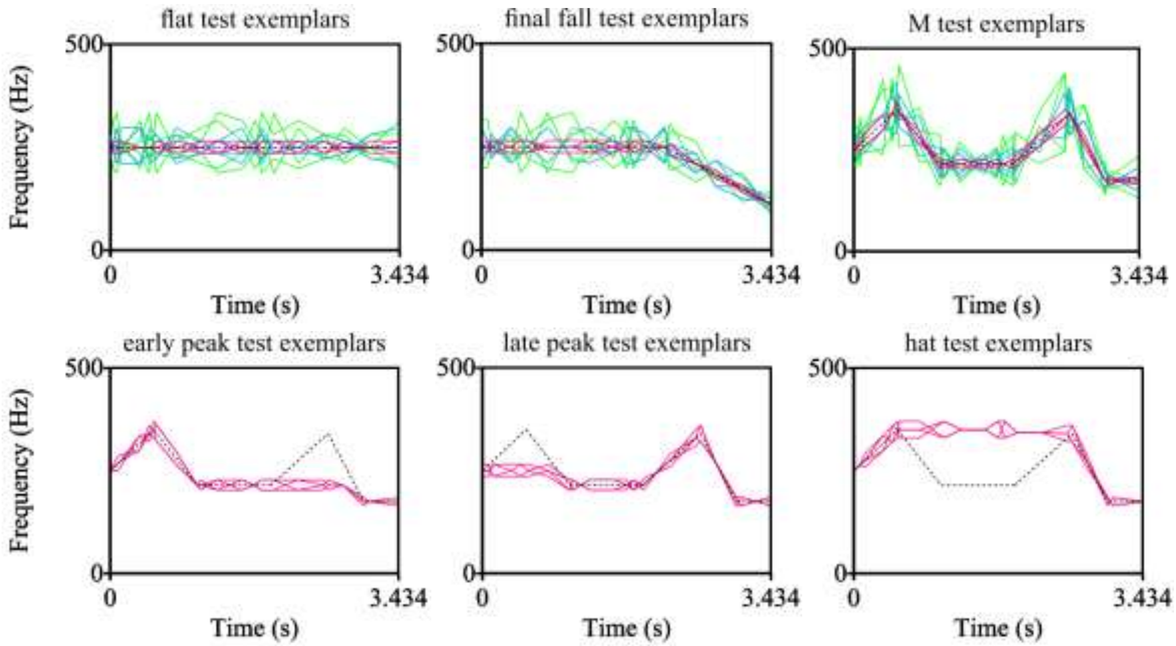
The stimuli consisted of intonation contours superimposed on a 15-syllable nonsense carrier phrase 3.43 seconds in length. The carrier was resynthesized from a recording of a sentence uttered by a male speaker of U.S. English obtained from the online Speech Accent Archive (Weinberger 2013). Resynthesis 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 preserved the rhythm characteristics of spoken English (see also Ramus & Mehler, 1999; Ramus, Nespor & Mehler, 1999; White, Mattys & Wiget, 2012).

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

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 four each mid- and high-level distortions wherein each syllable was perturbed by up to 3 or 5 semitones, respectively. 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. Example training and test items are shown in Figure 3.

Figure 3. 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.



Procedure

The experiment was administered individually 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 "talked like that". There were three such creatures, each of which was always paired with exemplars from the same set ('flat', 'final fall' or 'M').

The learning was passive, with items advancing automatically after an interval of 500ms. Following Carvalho & Goldstone (2012, 2014) and Xu & Tenenbaum (2007), training was blocked by category in order to avoid exclusive attention to features that distinguish the categories. Each block thus 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. The participants were aurally presented with the test items in random order, and their task was to categorize them my clicking on one of four 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 (Figure 4). The participants were instructed to click on this picture if the "sentence" they heard did not belong to any of the alien languages they had experienced. ("Is this *dax*, *wug*, *swinkie*, or one of these other guys?"). They responded by clicking on the appropriate picture. Item presentation advanced 500ms after registering a response. 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 4. The test response screen.

The dependent variable

The additional 'garbage bin' response option ("one of these other aliens" in Figure 4) allowed us to distinguish between *confusability*, $p(\text{Response}=\text{category}_i \mid \text{Response}\neq\text{'garbage'})$, and *acceptability*, $p(\text{Response}=\text{category}_i \mid \text{Response} = (\text{category}_i \text{ OR 'garbage'}))$ of the various category exemplars presented in the test stage. For example, if the 'M' contour is pronounced by the Wob in training, acceptability of test examples of 'M' and 'M' distractors is the frequency with which participants click on the Wob divided by the frequency with which they click on either the Wob or 'one of these other aliens'. Confusability of 'M' is the frequency with which the participants click on the Wob divided by the frequency with which they click on either the Wob, the Mup or the Biff.

We are primarily interested in acceptability rather than confusability based on both the design of the stimuli and our primary interest in category breadth. Increased category breadth is most purely indexed by increased acceptability (reduced probability of using the 'none of the above' option), because greater acceptability of low-confusability exemplars is difficult to account for by slower or poorer learning, being a relatively pure reflection of bias.

Broader categories are also expected to lead to increased confusability of the trained categories. However, the test stimuli were not designed in a way that would cause unacceptable stimuli to also be more confusable. A stimulus can be unacceptable without being confusable with other categories. While distractor and high-level prototype distortions are made to be less similar to the prototype than the training examples, they are not made to be more similar to the prototypes of the other categories, i.e. 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 no significant differences in confusability between the various stimulus types within any subject group.

Statistical 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). 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 the between-participant predictors: age and subject group (child vs. adult).

We used two operationalizations of age, both used within the child group only: continuous age in months and a binary split at the median (10;3). The results are unchanged across these two operationalizations. In the results section below, the output of regression analyses utilizes continuous age. Binary age is used *only* for the figures. Binary age is used for figures because there are few data points per response category per age, which makes scatterplots uninformative, as points form a small number of horizontal lines at the possible proportion values. Error bars in the graphs are 95% confidence intervals based on a proportion test (prop.test() in R, Newcombe 1998, Wilson 1927).

2.2 Results
For each type of test stimulus, we will report two statistical analyses, one comparing children to adults (group), and one examining the effect of age within the child group (children's age). In all statistical analyses reported, the dependent variables are binary (confusion and acceptance, as defined above). Age is entered as a continuous variable. However, the figures reported are based on a binary median split for age as discussed above.

Old and new low-level distortions
We begin by examining possible effects of group and children's age on responses to old and new low-level distortions. According to the "starting big" hypothesis (Arnon 2009, Arnon & Ramscar 2012, Tomasello 2003), children are expected to rely on memories for specific exemplars. Thus, the difference in acceptability between old and new exemplars at the same distance from the prototype is expected to be larger for children. The "starting small" hypothesis (Elman 1993, Newport 1990) does not presuppose reliance on memories for specific exemplars by either children or adults, thus it does not necessarily predict an interaction between stimulus type and age or subject group. If an interaction is found, it should be in the opposite direction from the predictions of the "starting big" hypothesis: if adults do rely on memories for specific exemplars to perform the categorization task, children are expected to not show as much reliance.

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). Thus, there is no exemplar effect for either group. There are also no significant interactions between subject group and stimulus type (b = -0.2904, $se(b)$ = 0.3853, z = -0.754, p = 0.4511 for confusability; b = 0.06762, $se(b)$ =

0.29233, $z = 0.231$, $p = 0.817$ for acceptability). There is likewise no interaction between age and stimulus type within the child group ($b = 0.2130$, $se(b) = 0.2028$, $z = 1.050$, $p = 0.2936$ for confusability; $b = -0.02532$, $se(b) = 0.20142$, $z = -0.126$, $p = 0.900$). These results are problematic for the "starting big" hypothesis, as neither group appears to show reliance on memories for specific exemplars and there is no expected decrease in reliance on exemplars with increases in children's age.[7]

Random high-level distortions

Higher-level (3 semitone and 5 semitone) distortions of the prototypes are one of the most direct ways to test the difference between "starting small" and "starting big". Both hypotheses predict interactions between group and children's age on the one hand and distance from the prototype on the other. However, the predicted interactions are in the opposite directions. The "starting small" hypothesis claims that children's categories are broader than those of adults. Therefore, we would expect children to show a smaller difference between low-level and high-level distortions, and young children to show an even smaller difference. The "starting big" hypothesis predicts that children's categories will be narrower than those of adults. Thus, we would expect children to show a larger difference between low-level and high-level distortions. Within the child group, we would expect younger children to show even less acceptance of high-level distortions. Both hypotheses predict the differences to occur mostly with acceptability, rather than confusability as the dependent variable, as the high-level distortions of a prototype are not necessarily closer to the prototypes of other categories.

The results are shown in Table 1. There is an effect of distance from the prototype: both 3-semitone and 5-semitone distortions are accepted less often than the 1-semitone distortions. There is no main effect of group as low-level distortions of the prototype are accepted approximately equally often by children and adults. There is a significant interaction between distance and group: children accept both 3-semitone and 5-semitone distortions more often than adults do. Table 2 shows that within the child group there is no main effect of age: low-level distortions of the prototype are accepted approximately equally often across the examined age range. There is a significant interaction between distance and age: rejection of high-level distortions increases with age. The direction of these interactions is predicted by the "starting small" hypothesis and is inconsistent with the "starting big" hypothesis: categories become narrower over the course of development.[8]

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

---

[7] There are no differences in confusability between new and old exemplars in either age group: ($b = 0.37$, $se(b) = 0.46$, $z = 0.81$, $p = .42$ for adults; $b = -0.46$, $se(b) = 0.40$, $z = -1.16$, $p = .24$ for children). The 'flat' category shows lower acceptability overall but there are no significant interactions with stimulus category.

[8] 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 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).

	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**

Table 2. The interaction between continuous age and distance from the prototype (where 1 semitone is 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.

	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*

The results so far are summarized visually in Figure 5, with the continuous effect of age represented categorically using a median split. The figure shows that previously experienced (TR1) and novel (1ST) exemplars at the same (low) distance from the prototype are accepted equally often by children and adults, showing no effect of long-term memory for specific exemplar contours. There is a clear effect of distance from the prototype, larger deviations being rejected more often than smaller deviations, and that the effect of distance from the prototype increases with age. Most strikingly, for the youngest children, 3-semitone distortions are accepted as often as 1-semitone distortions. Figure 5 also shows that the age range we examined likely shows the endpoint of the developmental trajectory: in agreement with Tahta et al. (1981), the older half of the children appear adultlike in their behavior in Figure 5, and are not statistically different from adults (Table 3).[9]

[9] 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. The point of Table 4 is simply that children that are older than 10;3 do not significantly differ from adult on our task (given our sample size of 20 children of this age and 40 adults).

Figure 5. 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.

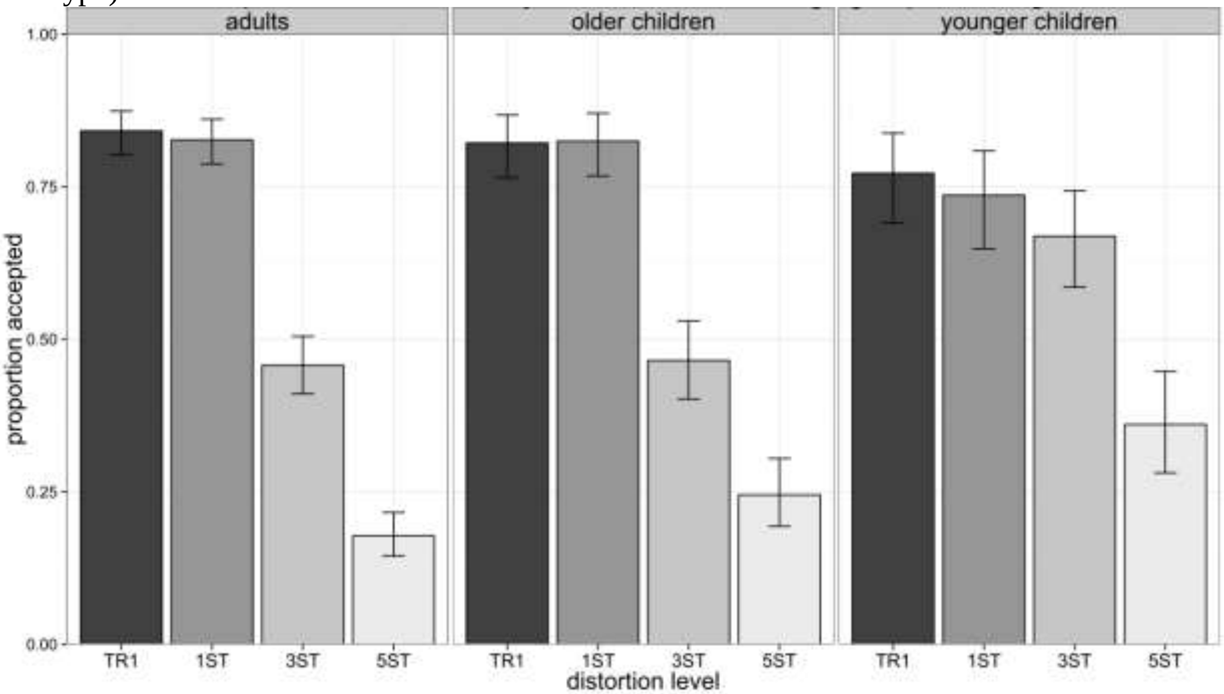


Table 3. The interaction between age (median split) and distance from the prototype: Only younger children significantly differ from adults in the effect of distance.[10]

	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***

[10] 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). Including the interaction with category makes it impossible to include a random effect for item -- due to there being few stimuli per combination of category and distortion level, the model does not converge -- which then makes the model less conservative in estimating the effects of age and subject group, for which the presented model has by-item random slopes.

Feature-removing distortions

We now turn to the 'M" distractors. These can be seen as high-level distortions of the 'M' prototype but -- unlike the distortions examined in the preceding section – they are not random. Rather, each distortion type eliminates one of the features characteristic of all of the training exemplars and the 'M' prototype. This manipulation provides us with another way to test whether children's categories are broader or narrower than those of adults. The "starting small" hypothesis predicts that children may pay attention to only some of the characteristic features of training examples, being unable to hold on to all of the features in working memory. They are therefore expected to accept test examples that share the attended features with the training examples while differing from them in the unattended features. The "starting big" hypothesis claims that children rely on memories of experienced instances and form narrower generalizations than adults. They should therefore be more likely to reject test examples that do not share all characteristic features with the training examples.

In addition to providing another way to distinguish between the two developmental hypotheses, these stimuli allow us to examine the relative importance of the characteristic features of a two-peaked intonation contour. We did not have any particular expectations about which particular features of a contour are most important for categorization.

The results are shown in Tables 4 and 5. Again, acceptability (classifying a contour as 'M' vs. 'garbage') is the dependent variable as the distractors are no more confusable than instances of 'M' are for either children or adults. Table 4 shows that children are more likely to accept /____ as an instance of /_/_ than adults are. Table 5 shows that acceptance of all distractors decreases with age within the child group. The results are again consistent with the "starting small" hypothesis: categories appear to narrow with age.

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.

	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

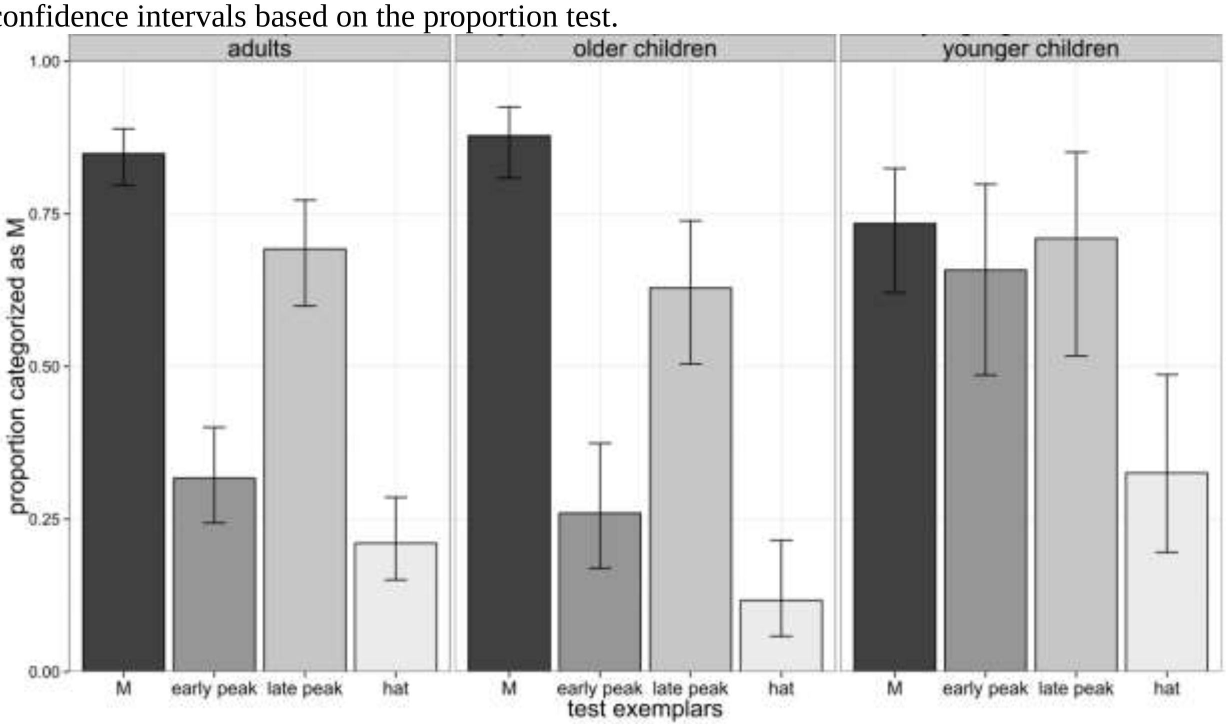
Table 5. The effect of continuous age for the 9- to 11-year-old children on acceptance of distractors lacking one of the features of 'M' (/_/_) as being instances of 'M'. Novel 1-semitone distortions of the 'M' prototype are the reference level to which the distractors are compared. Positive coefficients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

	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*

Figure 6 provides a graphical illustration of the results. There is a pronounced decline in acceptance of stimuli sharing only the first peak with previously encountered exemplars with age. Whereas adults reject such examples, younger children accept them as instances of ‘M’. Older children are once again adultlike in their behavior, suggesting that we are capturing the end of the developmental trajectory.

Figure 6. 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.



With respect to the relative importance of the various features of ‘M’, adults and older children appear to consider the second peak to be the most important feature of ‘M’ (acceptance rates are higher for ‘late peak’ stimuli compared to ‘early peak’ and ‘hat’ stimuli for adults: (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 stimuli are rejected slightly more often than early peak stimuli (b = -0.80, $se(b)$ = 0.38, z = -2.10, p = 0.036). Importantly, however, adults do pay attention to features other than the second peak, as ‘late peak’ distractors are accepted significantly less often than novel instances of ‘M’ by adults (b = -3.98, $se(b)$ = 0.86, z = -4.60, p < .0001).

In contrast, 9-10 year-old children accept stimuli that retain either the first or the second peak of the ‘M’ prototype. They are as likely to accept single-peaked distractors as novel two-peaked instances of ‘M’ (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). The hat distractor is still more likely to be rejected than the other stimuli (b = -2.67, $sd(b)$ = 0.83, z = -3.24, p = 0.001 for ‘M’ vs. hat): the continuously high pitch of the hat stimuli appears to make them very different from the previously experienced instances of ‘M’ and the single-peak distractors.[11]

Overall these results are consistent with the "starting small" hypothesis: children’s categories appear to be broader, as anything that shares one of the peaks with ‘M’ is in the ‘M’ category. However, the great importance of the second peak for adults was unexpected. and is discussed further in section 3.

3. Experiment 2: Compensation for declination

One possible reason for the greater salience of the second peak in the 'M' contour for adults is that new information tends to come at the end of the utterance (Clark & Clark 1978, Narasimhan & Dimroth 2008, Wundt 1900). Intonational peaks are often used to focus attention on new information. Thus, the end of the utterance is a more likely location for such a peak. Adults know this, while children may not. In fact, Narasimhan & Dimroth (2008) show that 3-5 year-olds have an ordering preference opposite to those of adults, placing new information first.

However, a less interesting explanation is that the second peak may have been perceptually larger, making its absence a greater deviation from the original stimuli. F0 usually declines throughout an utterance (e.g. Cohen & ‘t Hart 1967, Ladd 1984), likely for physiological reasons (Strik & Boves 1995). As shown by a number of experiments (e.g. Pierrehumbert 1979, Gussenhoven et al. 1997), listeners expect F0 declination and perceptually compensate for it. As a result of this perceptual compensation, the same F0 value is perceived as higher pitch when it occurs early rather than late in an utterance. We attempted to take this into account by making the second peak lower than the first peak by 0.5 semitones. However, it is possible that the amount of declination in our ‘M’ stimuli was insufficient. The aim of the present, follow-up experiment was therefore to evaluate whether the two peaks of the ‘M’ prototype were perceptually equal.

[11] There were few significant differences in confusability between the stimulus types for either age 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).

3.1. Methods

Participants
Thirty INSTITUTION undergraduates participated in the experiment in exchange for course credit. All were native speakers of U.S. English, and all reported normal hearing.

Stimuli
The M prototype was altered in one of two ways. For one set of stimuli, the second peak was either raised or lowered in 0.25 semitone increments, up to 3 semitones away from its original position. For the other set, the peaks were held constant but the magnitude of the final fall was progressively reduced by raising its endpoint in 0.25 semitone increments up to 5 semitones away from the prototype. Thus, there were 45 M-based stimuli: 12 with raised and 12 with lowered second peak, 20 with reduced final falls, and the original prototype. All contours were superimposed on the same /mimimi/ sequence used as a carrier in the main experiment. However, replicating Gussenhoven et al. (1997), the magnitude of the final fall had no effect on peak height perception as long as the height of the pitch peak was held constant. Thus, the present report abstracts away from this manipulation. All stimuli discussed in the following had the same F0 for the final fall endpoint as the M prototype.

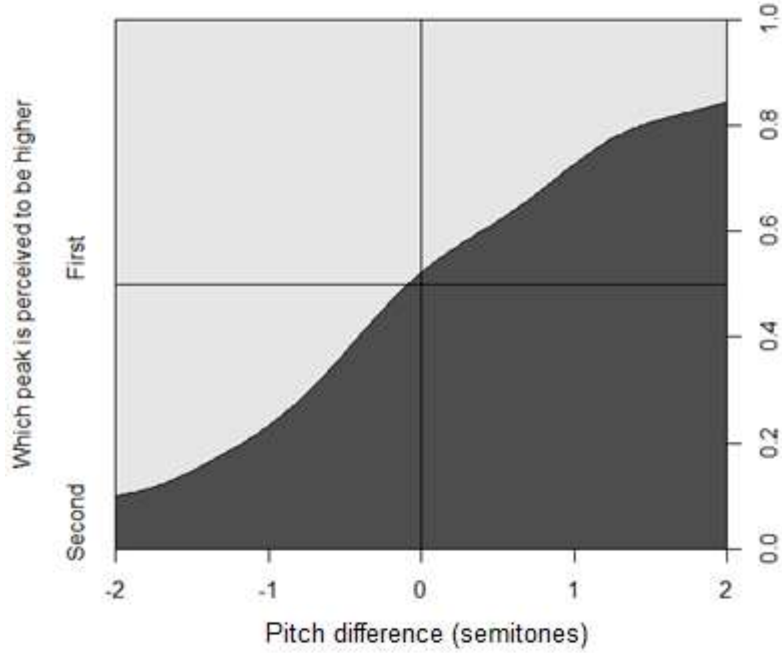
Procedure
Participants were tested individually in a quiet room. Presentation was blocked by type, with random trial and block order. The items were presented over headphones at a comfortable listening level. The task was to decide whether the second peak was higher or lower in pitch than the first (as in Gussenhoven et al. 1997). Responses consisted of clicking on the appropriate word on the screen (*higher/lower*). Additional visual cues were provided above each word by line drawings depicting the relevant contour type. Trials advanced 500ms after the system registered a response. Each item was presented once. Participants took approximately 5 minutes to complete the task.

3.2. Results and discussion

Figure 7 shows the results. For the difference in peak heights used in Experiment 1 (marked as 0 on the horizontal axis), the first peak was perceived as being higher approximately as often as the second peak (19 vs. 11, p = .22). For stimuli in which the height of the second peak was raised *or* lowered by .25 semitones from the original M prototype, each peak was perceived as the highest equally often (16/30, p = .86; 15/30, p = 1). A logistic regression model predicting response from F0 difference identifies 0.08 semitones above the difference used in Experiment 1 as the threshold above 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).[12] Thus the two peaks appear to be perceived as equally or almost equally high without training. In other words, compensation for declination was properly controlled in Experiment 1 (by making the second peak lower than the first by 0.5 semitones) and cannot explain why adults (and perhaps older children) found the second peak to be more important for categorizing contours into the ‘M’ category.

[12] .04 semitones / second * 2. While these judgments are on tones going up or down, ‘t Hart et al. (1990, p.30) argue that the judgments are made by comparing the initial and final frequencies of the tone.

Figure 7. Perceived relative heights of the two peaks of the 'M' contour as a function of the acoustic difference between them (in semitones, acoustic height of the second peak subtracted from the height of the first peak). The crossing lines indicate where the light/dark border would be if the second peak was judged as being higher than the first exactly 50% of the time with stimuli that are acoustically identical to the 'M' prototype from Experiment I (the ideal).



4. General discussion

In this paper we extended the study of category learning to the acquisition of intonation contours, temporally distributed fluctuations in a continuous parameter, pitch. To our knowledge, these kinds of stimuli have not been examined in either the categorization literature or the perceptual learning literature. We hypothesized that the great and greatly variable extent of intonation contours poses unique challenges to perception and categorization that should result in different categories being acquired by different learners. In this paper, we examined typically-developing native speakers of various ages. However, this work also opens an avenue to examine differences in categorization decisions during perceptual learning as a possible explanation of atypical native and second-language prosody.

We observed an age-related difference in categorization. Younger children appear to form broader categories than older children and adults when exposed to novel intonation contours in that they are more likely to accept large deviations from the prototype, whether those deviations involve random distortion or removal of characteristic features.

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). The greater breadth of children's intonation contour categories is in line with work on the development of concepts (Clark 1973, Mandler 2000, Pauen 2002, Rogers & McClelland 2004). It is also consistent with previous studies of categorization in which adults' judgments are shown to be based on attending to multiple dimensions while children's judgments are more likely to be based on a single dimension (T. Ward & Scott 1987. T. Ward et al. 1990, Thompson 1994, Visser & Raijmakers 2012), as well as with Schwarzer's (1997) finding that adults are more likely than children to categorize on the basis of feature values whose detection requires temporal integration. In spoken word recognition, young children have also 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. In syntax, 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 (though Linzen & Gallagher 2014 point out that the subject groups have not been equated on amount of training, and phonotactic generalizations are more general early on during the learning process).

Our findings are also consistent with relatively recent work on visual categorization that has questioned the global-to-local developmental shift (Cook & Odom 1992, Schwarzer 1997, Schwarzer et al. 1999, Thompson 1994, Thompson & Massaro 1989, Visser & Raijmakers 2012, T. Ward & Scott 1987, T. Ward et al. 1990, Wilkening & Lange 1989, Wills et al. 2013). In particular, Thompson (1994), T. Ward & Scott (1987) and T. Ward et al. (1990) showed that young children sorted objects primarily based on single dimensions rather than overall similarity, whereas adult sorts involved more dimensions, and that previous studies obscured this pattern by 1) averaging across subjects rather than analyzing individual data and 2) testing only on discovering the one dimension that is perfectly informative about category membership. When there are multiple redundant dimensions that are equally informative about category membership, children tend to focus on only one of the dimensions and memorize the few stimuli for which their chosen dimension does not work in predicting category membership (T. Ward & Scott 1987. T. Ward et al. 1990, Thompson 1994, Visser & Raijmakers 2012). Faced with the same task, adults are more likely to attend to multiple dimensions, consistently with the assumptions of the "starting small" hypothesis.

Previous work on categorization of temporally-extended patterns by children and adults is likewise consistent with "starting small". In their study on learning categories of haptically explored objects, Berger & Hatwell (1996) noted that

> "The exploratory movements used to process haptically object properties are specialized and adapted to gather information only on a restricted number of properties… Only few and low-cost exploratory movements are generally used by young children. This could facilitate selective attention and induce a differential weighting of one dimension. Accordingly, the fact that haptic perception tends to privilege a dimensional mode of response in young children is consistent with the assumption that the object is not completely processed early in age… Due to this specialization of haptic exploration, subjects could lose the global structure of the stimulus…" (pp.449-450)

Extending work on temporally extended stimuli to the domain of melody perception, Schwarzer (1997) also found that both children and adults tended to focus on single attributes of melodies to drive categorization decisions. However, adults were more likely to focus on attributes of the melody whose perception requires some degree of temporal integration, tempo and especially contour. On the other hand, 5-7 year-old children focused on loudness and timbre, which could be assessed at any point during the melody and therefore did not require temporal integration. For example, in Experiment 2, 83% of the 5-6 year-olds focused on loudness, compared to 8% of adults; on the other hand, 52% of adults focused on melodic contour (going up, down or up and down), compared to 0% of children. Again, this result is consistent with the "starting small" hypothesis: adults are able to easily integrate the separate tones of a melody into a contour, while children are not.

Possible reasons for children having broader categories

Why are children so broad-minded? One might be tempted to propose that children are simply more accepting than adults across the board, irrespective of stimulus properties. However, several findings from this study as well as related studies argue against this interpretation. First, the developmental change is much stronger for the early peak stimuli than for the late peak or hat stimuli (Figure 6). Second, the change is not observed at all for low-level distortions (Figure 5). Third, the same subjects also participated in a study in which they learned a miniature artificial language featuring plural markers and were asked (among other things) to judge singular-plural pairs of words like *blup-blutʃa* ("is this the right plural for this singular?"). In that experiment, the child participants were no more accepting than the adults (though the groups were not close to ceiling or floor). Finally, Rabi & Minda (2014b, Experiment 2), who trained children and adults on high distortions of a visual (dot pattern) prototype, found that children were *less* likely to accept distortions of the prototype than adults were. Thus, we suggest that the greater acceptance of deviations observed in the present study is not simply due to children being likely to accept anything. Rather, they accept distortions of the prototype that still share many features with it when trained on stimuli that pose working memory challenges.

A possible explanation for the present results is thus the fact that children have a limited working memory (Newport 1990). Working memory limitations can lead to greater acceptance of deviations from previously experienced exemplars in multiple ways. First, children's memories of the experienced exemplars may be more "blurry", and thus more underspecified, than adult memories because features are missed during encoding (Berger & Hatwell 1996, Charles-Luce & Luce 1990, 1995, Fikkert & Levelt 2008). Second, some features of stored exemplars may not be accessed when the test stimuli are to be compared with stored exemplars (Fennell & Werker 2003, Werker & Fennell 2009). Third, features of the test stimuli may be missed. At the decisional level, knowing that their memory is limited, children may be more accepting of mismatches between current perception and memory. Future work may test this hypothesis by determining if categorization breadth correlates with working memory capacity across individuals or is increased when participants are placed under a working memory load (as in Wills et al. 2013).

Differences in processing speed between children and adults may also contribute to broader categorization in children. 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, T. Ward & Scott 1987, Wills et al. 2013), and chidren's processing speed is slower, then children may prefer to respond on the basis of a single dimension to conserve time and reduce effort that would have been much higher than for adults if both groups took into account the same number of dimensions. Under this hypothesis, category breadth should increase when participants are asked to respond quickly.

A related possibility is that children's perceptual abilities are underdeveloped compared to those of adults. This hypothesis was first raised by Gibson & Gibson (1955) and tested for visual perception by Ashkenasy & Odom (1982), Odom & Cook (1984), and Thompson & Massaro (1989). As summarized by Thompson (1994, p.1643), children may "not be as capable as adults at distinguishing slight differences in sizes and brightnesses of objects… In their attempt to attend selectively to a single, separable dimension, children may perceive identical matches of features that are actually slightly different." Again, the 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 (the high-level distortions in Figure 5). Future work may test this explanation by replicating Experiment 2 with children. If this hypothesis is true, the slope of the line in Figure 7 should be shallower in children (e.g. Thompson & Massaro 1989). Alternatively, adults may differ from children in assigning a special value to identity (as opposed to mere similarity) on the decisional level (L. Smith 1989), so that novel stimuli have to match the experienced examples of a category exactly on the attended dimensions to be placed into the category.

Another decisional-level 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" (Xu & Tenenbaum 2007; see also Jenkins et al. 2014). This may make them more likely to cover the perceptual space of possible contours with the categories they have. Since this hypothesis locates the difference between children and adults on the decisional level, children and adults should perceive and store contours equally well, which could be assessed using old/new recognition, discrimination, and delayed repetition tasks. Because this explanation assigns credit for the difference between the age groups to linguistic experience, it also predicts that age-related changes in category breadth may be more likely to be found with linguistic rather than non-linguistic stimuli of comparable length and complexity (e.g. melodies in Schwarzer 1997).

Applicability of exemplar models to intonation

It appears that both children and adults are fairly broad-minded in their treatment of intonation contours in that neither group appears to rely on their memory of 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 5.[13] This result is inconsistent with the assumptions of the "starting big" hypothesis (Arnon 2009, Arnon & Ramscar 2012, Tomasello 2003) and contrasts with classic findings for simple visual patterns (Medin & Schaffer 1978, Nosofsky 1988, Posner & Keele 1968, J. Smith & Minda 1998). The lack of a difference between old and new exemplars suggests a relatively abstract representation of intonation patterns, abstract enough for the new exemplars not to differ representationally 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 categorization literature are the 8-dimensional stimuli seen

[13] See also Cristia & Peperkamp (2012), Linzen & Gallagher (2014) for the same finding with perception of novel vs. old segmental sequences drawn from the same general class. See Kapatsinski (2014) for other differences between the results of studies of linguistic and visual categorization.

to disfavor exemplar memorization in Minda & J. Smith (2001). In real life, contours can also vary in duration and the number of possible inflection points and perceptibly different pitch levels for those is even higher (e.g. ‘t Hart et al. 1990).

In addition, the dimensions defining a contour are redundant (as in the category structures examined by T. Ward & Scott 1987 *et seq*). For example, all training exemplars of ‘M’ have two peaks and a valley in between, features absent in exemplars of the other contour categories. As a result, the categories of contours are well separated: the contrast between contour categories does not rely on the accurate perception of a single feature that would be easily obscured by noise, lapses of attention or motor variability. As suggested by J. Smith et al. (1997) and J. Smith & Minda (1998), well separated categories are less likely to exhibit exemplar effects than the poorly separated categories used in much previous work on categorization.[14] Work by T. Ward & Scott (1987), T. Ward et al. (1990), Thompson (1994), Schwarzer (1997) and Wills et al. (2013) suggests that both children and adults exposed to such categories pick a subset of features to identify the category by, although different learners pick different features (see also Idemaru et al. 2012 for phonetic categories). The fact that the features are redundant allows different learners to focus on different features – depending on differences in cognitive style, perceptual abilities and prior experience – without jeopardizing the ability of different learners to converge on the same categorization.

We would argue that it is typical for linguistic features to be highly redundant (e.g. Hockett 1965, Kapatsinski 2014, Pierrehumbert 2000 see also T. Ward et al. 1990 for other categories) and thus for linguistic categories to be well separated (e.g. Baese-Berk & Goldrick 2009, Liljencrants & Lindblom 1972, Pierrehumbert 2001, Stevens & Keyser 1989). A few reasons for this state of affairs are mentioned by Pierrehumbert (2000), who writes:

> “it is worthwhile to bear in mind how sparsely languages in general sample the cross-product of available units. Because of phonotactic constraints, most combinations of phonemes do not represent possible words and most phonotactically possible words do not happen to be real words. Syntactic and semantic constraints have the consequence that most combinations of real words do not constitute potentially observable sentences. Similarly, we need to work out what factors cause gaps in the set of intonation patterns observed. Are there phonotactic factors that are not yet understood? Are phrasal intonation patterns lexicalized as single units, with some being accidentally missing? Are some potential patterns missing because the meanings of the component parts cannot be coherently combined?” (Pierrehumbert (2000, p.27; see also Kapatsinski, 2014).

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 attributes (e.g. Wright et al. 2000, Olsson & Poom 2005). The working memory difficulties involved in memorizing a contour exemplar may be especially severe in that a contour consists of a sequence of differing values of a single feature, which may result in mutual interference between different pitch levels in the sequence (see, e.g.

[14] Though this factor is not seen to disfavor exemplar memorization in Minda & Smith (2001), that study may have used categories that are small enough for all exemplars to be memorized despite good category separation.

Logie et al. 1990, Lustig et al. 2001, Obenauer & Kliegl 2006, for the role of interference between similar stimuli in limiting working memory).

The category structure for natural categories of intonation contours, as well as other phonological units, also 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), in which there are no features that are individually necessary and jointly sufficient for category membership. However, natural categories or phonological forms are well-defined (Kapatsinski 2014). Thus, in the absence of noise, mismatch in a single feature or letter is sufficient to eliminate repetition priming (e.g., if correctly perceived, the pseudoword *blick* would 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 intended to be produced by the speaker/writer for the word to be recognized.[15] Similarly, 't Hart et al. (1990: 84) describe 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), as the characteristic features of the contour category are very informative with respect to category membership. Language learners may come to know this fact about categories of language forms, and expect even novel categories of contours to be structured in this way.

Finally, the training regime differs from that in the classic studies of categorization in that we did not train participants to criterion on the training exemplars. Training and testing on the same exemplars to criterion may have led the participants in Posner & Keele (1968) *et seq* to focus on remembering the specific exemplars presented and reduced generalization to novel exemplars. In the present study, participants received a fixed amount of training. Furthermore, the amount of training was fairly limited, perhaps, too limited to form lasting memories of specific exemplars of the contours. J. Smith & Minda (1998) provide evidence for reduced exemplar memory effects in the early stages of training. For phonological units, gradual development of memory for specific experienced structures accompanied by rapid sensitivity to broad generalizations is shown by Linzen & Gallagher (2014).

Is less more?

Is being broad-minded helpful for language acquisition? 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 be more when it comes to language acquisition (Elman 1993, Newport 1990). Rogers & McClelland (2004, p.96) argue that "there may be good computational reasons [for a neural network] … to begin with … undifferentiated internal representations [i.e., a single category]. Specifically, such an initial state would permit very rapid learning about properties common to all things", or in our case, all intonation contours (see also Griffiths et al. 2007, Love et al. 2004). There may be more need to learn what intonation contours in general are like for children than for adults. Hence, the developmental timecourse may actually present children and adults with a bias that is optimal for their state of knowledge.

[15] As S. Armstrong et al. (1983) and Wierzbicka (1990) argue, features can be necessary despite instances of their values being difficult to identify. For instance, Wierzbicka (1990) claims "made to sit on" to be a necessary feature of a chair. While it might be difficult to identify whether a given object is made to sit on, this does not make the feature unnecessary.

Broader categories may also be helpful for dealing with variability in production and perception (e.g. Lively et al. 1993, Logan et al. 1991). Again, the ability to deal with variability may be especially helpful for children, as children have less experience with language and thus more likely to encounter input that they have not encountered before or input that is not even particularly similar to previous experience. The child's own productions, which form a significant part of his/her experience with language, are also more variable than those of adults (e.g. Kent & Forner 1980, Lee et al. 1999). Kapatsinski (2013, 2014) argues that starting with general categories of phonological structures is helpful for dealing with novel inputs, and that this ability is especially important when one's lexicon is small. Kapatsinski (2014) illustrates the point by examining a specific-to-general rule learning model called the Minimal Generalization Learner (Albright and Hayes 2003). The rules in this model start out very specific, particular to individual words, and only gradually become more abstract. For example, given the word pairs *bank-banked* and *link-linked*, the model would generalize the rule 0→ed / X[-back;+syl]nk___ where the feature specification after the slash defines a category of words to which the 0→ed rule applies.[16] If the model is then presented with the verb *lick*, which it has never encountered, the model would have no idea what the past tense of the verb is: it knows of no rule applicable to this word, as *lick* does not fall into any of the word categories associated with rules. In order to deal with novel inputs, potentially very different from prior experience, a model of phonology needs to extrapolate beyond the input. Similarly, with intonation contours, it may be helpful to extrapolate beyond the input so that the same meaning could be activated by very different realizations of a contour that do not provide a good match to prior experience.

We should note that the findings providing most direct support for the "starting big" hypothesis, come from production (e.g. no judgment, preferential looking, act-out or "weird word order" studies providing evidence for specific-before-general acquisition order are mentioned in the extensive review by Ambridge & Lieven 2011). Given that all of our data come from perception, we do not have any data regarding the appropriateness of the "starting big" timecourse for *production* of intonation contours. It may well be the case that despite accepting a wide variety of contours as belonging to an intonational category, children may tend to re-produce only the contours they have experienced, and may even be more likely to do so than adults (Tahta et al. 1981). In a broader context, more exact *immediate* repetitions of intonation contours on the parts of children are unsurprising given that children have been observed to exhibit more priming (e.g. Castles et al. 2007, Farrar 1992, Rubino & Pine 1998, Stanovich 1980, West & Stanovich 1978). However, if better repetition of contours is observed with a delay, we may have evidence that the greater likelihood of children to accept distractors and high-level distortions is a bias that affects categorization decisions rather than a difference in perceptual representations of category members (cf. Ashby & Perrin 1988, Schwarzer 1997).

Children's ability to reproduce intonation contours in their full phonetic detail would be a welcome result from the "less is more" standpoint, as it is decisional breadth -- rather than inability to store perceptual detail -- that is potentially useful for language acquisition (see also Werker & Fennell 2009). As discussed above, it may be helpful for children to form broad categories that can be associated with various properties and then easily accommodate novel percepts. Given that natural-language categories are characterized by many redundant features, these novel percepts are likely to fall into the right category even if the categories are broader

[16] Note that linguistic rules are neither unidimensional nor verbalizable, unlike rules in the categorization literaure (e.g. Smith et al. 2008). Rather, rules are changes in context, where the context is defined by a structure consisting of necessary and sufficient features.

than those of adults. Given broad categories, one can quickly learn the properties of many possible percepts that fall into the category by associating the category with these properties. However, given the need to speak as well as understand, it is not helpful to store only some properties of a contour, as *all* characteristic properties of the contour need to be reproduced for an adult listener to perceive it as the same contour (see also ‘t Hart et al. 1990, p.84).

Final prominence

A final, unexpected finding was that the late peak was more important than the early peak for /_/_category membership for adults whereas children appeared to weight both features equally. A follow-up experiment ruled out a simple perceptual explanation for this result: both peaks are judged approximately equally high, thus their absence should be equally salient. This trajectory from (near-)equal weighting of features to selective weighting of features that are particularly informative has been proposed by L. Smith (1989) for object recognition in vision as well as for speech sound perception (e.g. Pisoni et al. 1994). Our results thus provide support for this hypothesis. However, note that L. Smith (1989) proposes 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 adults weight the second peak more than the first peak, they also judge stimuli lacking both peaks to be instances of ‘M’ less often than stimuli that have both peaks. In contrast, children weight the two peaks equally but accept stimuli having either peak as instances of ‘M’ as often as they accept stimuli having both peaks.

Why is it the second peak that is singled out as particularly important by adults? We suggest that 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) finding that 3-5 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 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.

5. Conclusion

The present study extended work on perceptual learning to temporally extended, suprasegmental patterns, namely intonation contours. We documented differences in categorization relating to category breadth between children and adults. 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 are equidistant from the prototype. However, the categories they formed did differ. Given the same experience, 9-10 year-old children formed broader categories than adults, being more likely to accept major distortions of the prototype into the category. This included even distortions that removed some of the features that were present in all training examples. This gradual narrowing is predicted by maturational improvements in working memory (Newport 1990), which allows adults to pay attention to all features characterizing a contour and integrate them into a unified whole. It is also predicted by increasing language experience: as children mature, their contour categories grow in number, potentially changing their estimated probability of encountering a new intonation contour (cf. Xu & Tenenbaum 2007 for word meanings). One way to distinguish between these hypotheses would be to examine bilinguals, who are exposed to even more intonation contour categories than age-matched monolinguals. In addition, adults were found to pay more attention to the end than to the beginning of the contour for the purposes of categorization. Attention to the end may be due to the end of the sentence being the likely locus of intonational prominence and the likely location for the 'root', the most informative part of an intonational contour ('t Hart et al. 1990). Individual differences in categorization affecting perceptual learning of prosody are a possible explanation for atypical prosody observed in populations with atypical processing styles, e.g. children with Autism Spectrum Disorders, or prior linguistic experience (e.g. second language learners of an intonation language with a tonal first language background).

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 MAR. The content is solely the responsibility of the authors and does not necessarily reflect the views of NICHD.

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational / experimental study. *Cognition*, *90*(2), 119-161.
Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, *40*(1), 41-61.
Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.
Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529-55.
Armstrong, L. E., & Ward, I. C. (1926). *A handbook of English intonation*. Leipzig: Teubner.
Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 263–308.
Arnon, I. (2009). Starting big. The role of multi-word phrases in language learning and use. Ph.D. Thesis, Stanford University.
Arnon, I., & Ramscar, M. J. A.(2012) Granularity and the acquisition of grammatical gender: How order of acquisition affects what gets learned. *Cognition*, *122*, 292-305.
Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124-150.
Aschkenasy, J. R., & Odom, R. D. (1982). Classification and perceptual development: Exploring issues about integrality and differential sensitivity. *Journal of Experimental Child Psychology, 34*, 435-448.
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.
Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527-554.
Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178-98.
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.
Bent, T. (2005). Perception and production of non-native prosodic categories. Ph.D. Thesis, Northwestern University.
Berger, C., & Hatwell, Y. (1993). Dimensional and overall similarity classifications in haptics: A developmental study. *Cognitive Development*, *8*(4), 495-516.
Berger, C., & Hatwell, Y. (1995). Development of dimensional vs. global processing in haptics: The perceptual and decisional determinants of classification skills. *British Journal of Developmental Psychology*, *13*(2), 143-162.
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.
Bolinger, D. (1964). Intonation as a universal. In *Proceedings of the 5th Congress of Phonetics, Cambridge 1962* (pp. 833-848).
Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.
Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
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.
Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*, 205-215.
Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*, 727-735.
Cho, H. (2010). A weighted-constraint model of F0 movements. Ph.D. Thesis, MIT.

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. (1973). What's in a word? On the child's acquisition of semantics in his first language. In: T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-110). New York: Academic Press.
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.
Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities the auditory scaffolding hypothesis. *Current Directions in Psychological Science*, *18*(5), 275-279.
Cook, G. L., & Odom, R. D. (1992). Perception of multidimensional stimuli: A differential-sensitivity account of cognitive processing and development. *Journal of Experimental Child Psychology*, *54*(2), 213-249.
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.
Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*, 144-152.
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.
Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.
Falter, C. M., Elliott, M. A., & Bailey, A. J. 2012. Enhanced visual temporal resolution in autism spectrum disorders. *PLoS One*, *7(3)*, e32774.
Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental psychology*, *28*(1), 90-98.
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.
Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, *30*(3), 321-338.
Fikkert, P., & Levelt, C. (2008). How does Place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in phonology: theory, perception, and acquisition* (pp.231-68). Berlin: Mouton de Gruyter.
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.
Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Blackwell.
Frith, U., & Happe, F. (1994). Autism: beyond 'theory of mind'. *Cognition*, *50*, 115–32.

Fujisaki, H. (1983). Characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (Ed.), *The production of speech* (pp.37-47). New York: Springer.

Garner, W. R. (1974). *The processing of information and structure*. New York: Erlbaum.

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.

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

Gorisch, J., Wells, B., & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech*, *55*, 57-76.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*, 375-401.

Griffiths, T. L., Canini, K. R., Sanborn, A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*, 323–328.

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.

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

Haywood, N. R., I-C. J. Chang, & V. Ciocca. (2011). Perceived tonal continuity through two noise bursts separated by silence. *Journal of the Acoustical Society of America*, *130*, 1503-14.

Hedberg, N., Sosa, J. M., & Fadden, L. (2006). Tonal constituents and meanings of yes-no questions in American English. *Speech Prosody 2006*.

Hockett, C. F. (1965). Sound change. *Language*, *41*, 185-204.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366-415.

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

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950-3964.

Jenkins, G. W., Samuelsson, L. K., Smith, J. R., & Spencer, J. P. (2014). Non-Bayesian noun generalization in 3- to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science*, *EarlyView*. 10.1111/cogs.12135

Jones, D. (1957). *An outline of English phonetics*, 8th ed. Cambridge: Heffer.

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

Kapatsinski, V. (2013). Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language*, *89*(1), 110-148.

Kapatsinski, V., & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. A. Moravcik, H. Ouali & K. M. Wheatley (Eds.), *Formulaic*

language. *Vol. 2: Acquisition, loss, psychological reality and functional explanations* (pp.499-520). Amsterdam: John Benjamins.
Kemler Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning & Verbal Behavior*, *23*, 734–759.
Kemler Nelson, D. G. (1988). When category learning is holistic: A reply to Ward and Scott. *Memory & Cognition*, *16*, 79-84.
Kent, R.D., & Forner, L.L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, *8*, 157-168.
Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.
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.
Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*, 1455-1468.
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.
Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 447-454.
Liberman, M. (1978). The intonational system of English. Ph.D. Thesis, MIT.
Liberman, M., & Sag, I. (1974). Prosodic form and discourse function. *Chicago Linguistic Society*, *10*, 416-27.
Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839-862.
Linzen, T., & Gallagher, G. (2014). The timecourse of generalization in phonotactic learning. *Proceedings of the Annual Meetings on Phonology*, *1*, 1-12.
Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.
Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.
Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*(1), 55-74.
Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: A network model of category learning. *Psychological Review, 111*, 309–322.
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.
Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199-207.

Maddox, W., & Chandrasekaran, B. (2013). Tests of a dual-system model of speech category learning. *Bilingualism: Language and Cognition*, *FirstView*, 1-20. doi:10.1017/S1366728913000783

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.

Makarova, V. (2001). Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics*, *29*, 137-54.

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.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, *42*, 682-702.

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.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207-238.

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.

Mishra, T. (2008). Decomposition of fundamental frequency contours in the general superpositional intonation model. Ph.D. Dissertation, Oregon Health & Science University.

Mitchell, T. M. (1980). The need for biases in learning generalizations. Rutgers University Technical Report: CBM-TR-117.

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.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Nakano, T., Ota, H., Kato, N., & Kitazawa, S. (2010). Deficit in visual temporal integration in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1027-30.

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

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine, 17*, 873–890.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11-28.

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

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54-65.

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

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601-626.

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. (1982). Perceptual sensitivity, integral perception, and similarity classifications of preschool children and adults. *Developmental Psychology*, *20*, 560-567.

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.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107-123.

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.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*, 205-20.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*, 1015-1028.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

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

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. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, *16*(5-6), 691-698.

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., & Levi, S. V. (2007). Some observations on representations and representational specificity in speech perception and spoken word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*, 3-18. Oxford University 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

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. (2104b). Perceptual category learning: Similarity and differences between children and adults. *Proceedings of the Annual Conference of the Cognitive Science Society*, *36*.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512-521.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

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., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 75-86.

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.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, *133*(1), 63-82.

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.

Rubino, R. B., & Pine, J. M. (1998). Subject–verb agreement in Brazilian Portuguese: what low error rates hide. *Journal of Child Language*, *25*, 35-59.

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

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022-2039.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*(4), 425-443.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language*, *24*(2), 210-231.

Schwarzer, G. (1997). Analytic and holistic modes in the development of melody perception. *Psychology of Music*, *25*(1), 35-56.
Schwarzer, G., Küfer, I., & Wilkening, F. (1999). Learning categories by touch: On the development of holistic and analytic processing. *Memory & Cognition*, *27*(5), 868-877.
Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, *25*(4), 425-437.
Shakespeare, W. (1603). *The tragicall historie of Hamlet, Prince of Denmarke*. London: John Trundell.
Shepp, B. E., Burns, B., & McDonough, D. (1982). The relation of stimulus structure to perceptual and cognitive development: Further tests of a separability hypothesis. In J. Becker, F. Wilkening & T. Trabasso (Eds.), *Information integration by children*. Hillsdale, NJ: Erlbaum.
Shport, I. A. (2011). *Cross-linguistic perception and learning of Japanese lexical prosody by English listeners* (Doctoral dissertation, University of Oregon).
Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*, 1097-1115.
Smith, J. D., & Kemler, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*, 137–159.
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, J. D., Murray Jr, M. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 659-80.
Smith, J. D., & Shapiro, J. H. (1989). The occurrence of holistic categorization. *Journal of Memory & Language*, *28*, 386–399.
Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*(1), 125-144.
Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology*, *10*(4), 502-532.
Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, *24*(2), 279-298.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381-382.
Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32-71.
Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, *65(1)*, 81-106.
Strik, H., & Boves, L. (1995). Downtrend in f_0 and P_{sb}. *Journal of Phonetics*, *23*, 203-20.
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.

Thorsen, N. (1978). An acoustical analysis of Danish intonation. *Journal of Phonetics*, *6*, 151-75.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

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).

van der Zande, P., Jesse, A., & Cutler, A. (2013). Lexically guided retuning of visual phonetic categories. *The Journal of the Acoustical Society of America*, *134*(1), 562-571.

van Santen, J., Kain, A., Klabbers, E., & Mishra, T. (2005). Synthesis of prosody using multi-level unit sequences. *Speech Communication*, *46*(3), 365-375.

Vinnik, E., P. M. Itskov, & E. Balaban. (2012). β- And γ-band EEG power predicts illusory auditory continuity perception. *Journal of Neurophysiology*, *108*, 2717-24.

Visser, I., & Raijmakers, M. E. (2012). Developing representations of compound stimuli. *Frontiers in psychology*, *3*.

Ward, N. G. (2014). Automatic discovery of simply-composable prosodic elements. *Speech Prosody 2014*, pp 915-919.

Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 1125-1138.

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.

Watson, D., Tanenhaus, M. K., & Gunlogson, C. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science*, *32*, 1232-1244.

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

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.

Werker, J., & Fennell, C. T. (2009). Infant speech perception and later language acquisition: Methodological underpinnings. In J. Colombo, P. McCardle, & L. Freund. (Eds.), *Infant pathways to language: Methods, models, and research directions*. New York: Psychology Press.

West, R. F., & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development*, *49*, 717-727.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665-679.

Wierzbicka, A. (1990). Prototypes save. In S. L. Tzohatzidis (Ed.), *Meanings and prototypes: Studies in linguistic categorization*. Routledge.
Wilkening, F., & Lange, K. (1989). When is children's perception holistic? Goals and styles in processing multidimensional stimuli. In T. Gloherson & T. Zelniker (Eds.), *Cognitive styla and cognitive development* (pp.141-173). Norwood, NJ: Ablex.
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.
Yang, C. (2011). *The acquisition of Mandarin prosody by American learners of Chinese as a foreign language*. Ph.D. Thesis, The Ohio State University.
Yang, C., & M. K. M. Chang. (2010). The perception of Mandarin Chinese tones and intonation by American learners. *Journal of the Chinese Language Teachers Association*, *45*, 7-36.