The Issue of Methodology in Prosodic Development

Jill C. Thorson\textsuperscript{a}, James L. Morgan\textsuperscript{b} and Rupal Patel\textsuperscript{a}

\textsuperscript{a}Northeastern University; \textsuperscript{b}Brown University

j.thorson@neu.edu

Age, attention, memory, as well as speech and cognitive abilities of a participant all impact experimental design choices. These considerations are perhaps even more important when the population is early or atypical in development. Many researchers have debated when and to what extent children can understand varying prosodic contours and distinctions, with disputes surrounding possible asymmetries between perception, production, and comprehension ([1], [2]). Interestingly, underlying many of these arguments is methodology, with issues such as level of cognitive demand, working memory requirements, and complexity and naturalness of the task affecting outcomes. My research addresses some of these issues and explores several paradigms to test prosodic knowledge over development, with both typically- and atypically-developing children.

In one line of research, I ask how intonation and information structure affect early perceptual abilities (Experiments 1 and 2) and how they are employed during early production (Experiment 3) ([3], [4]). Experiments 1 and 2 used an eye-tracking paradigm, providing a more in depth analysis into the processing patterns of young toddlers. Both experiments also employed a discourse-based procedure that presented children with words embedded in continuous speech rather than isolation (Figure 1). Experiment 1 demonstrated that either newness or the presence of a pitch accent guided 18-month-olds’ attention during reference resolution. Experiment 2 showed that a complex pitch accent and contrastiveness aid 24-month-olds in learning a novel word.

Experiment 3 examined the intonation of toddlers and adult speakers of American English during a spontaneous speech task. Using a novel interactive game, target nouns were produced as one of three types: new, given, or contrastive (Figure 2). This study showed how toddlers phonologically and acoustic-phonetically produced information structure differences in a manner comparable to the ones found in adult speech. The aim of using this method was to create an ecologically valid scenario that would encourage children to speak in a more naturalistic manner. Previous work using a corpus-based analysis of at-home recorded spontaneous speech has also shown that toddlers approximate adult intonation patterns from roughly the onset of speech ([4], [5]).

A second line of research uses a novel methodology called Prosodic Marionette to test prosodic knowledge in populations that may have speech difficulties due to age or physical impairment. Prosodic Marionette is a program that allows a user to physically manipulate an utterance on a computer screen (Figure 3, [6]). Words can move up/down to create changes in pitch, longer/shorter to change duration, and closer/farther apart to alter rhythmic patterns. The tail at the end allows the user to change the final boundary tone. This paradigm has been shown to be successful for use with 4-, 7- and 11-year olds, and current work is using it to study the prosodic abilities of populations with dysarthria due to Cerebral Palsy or ALS.

These studies take extra care to use paradigms and novel procedures that are more appropriate for the populations analyzed. This ensures a better and more accurate understanding of how young children typically employ different aspects of the prosodic system. (Word Count: 497)

\textsuperscript{*}The first three studies discussed are in collaboration with Dr. James Morgan at Brown University. The second two line of research are with Dr. Rupal Patel at Northeastern University.
**Figure 1.** Experimental trials from an eye-tracking paradigm for Experiments 1 and 2 testing early attention and word learning abilities in toddlers at 18- and 24-months-old.

**Figure 2.** Experimental set-up for a production study that elicits target words in an ecologically valid and spontaneous interaction. Children are asked to help the experimenter place the characters and their items on the spaceship based on a map that they are given.

**Figure 3.** Prosodic Marionette screenshot, an innovative visuo-spatial program that allows the user to physically manipulate the elements of an utterance. Items can be moved up/down (pitch), side to side (rhythm/pauses), and elongated (duration). Final boundary tones can also be manipulated up and down ([6]).
References
The role of prosody in five-month-old infants’ language discrimination
Laurence White¹, Claire Delle Luche² and Caroline Floccia¹

¹Plymouth University, ²Essex University

laurence.white@plymouth.ac.uk

Young infants pay particular attention to the prosodic features of spoken utterances, which may serve to promote interaction and also to bootstrap word segmentation and hence acquisition of vocabulary and syntax (e.g., Johnson & Jusczyk, 2001). This perceptual bias naturally entails sensitivity to prosodic contrasts between languages; indeed, patterns of language discrimination by infant listeners have been taken as support for perceptually-based “rhythm” classes, despite the weak phonetic evidence for such typological distinctions (e.g., Dauer, 1983). Furthermore, it has been suggested that this apparent categorical rhythm setting has important implications for the nature of the primary units – moras, syllables, stress-delimited feet – of the language learner’s segmentation of the speech stream (Cutler, 1994).

In previous studies, young infants have been shown to discriminate languages from different putative rhythm classes, but not languages within classes, except in the case that one or both languages is familiar to the infant. Thus, for example, English infants do not distinguish “syllable-timed” Italian and Spanish nor “stress-timed” Dutch and German (Nazzi, Jusczyk & Johnson, 2000), but Spanish infants can distinguish “syllable-timed” Catalan and Spanish (Bosch & Sebastián-Gallés, 2001).

The testing of adult sensitivity to prosodic distinctions is compromised by their phonemic awareness, which makes distinguishing languages a comparatively straightforward task. Thus, adult discrimination experiments have exposed listeners to stimuli modified to eliminate segmental information and focus solely on timing (e.g., “sasasa speech”). Initial results suggested that adult discrimination was also governed by categorical rhythm class distinctions (e.g., Ramus, Dupoux & Mehler, 2003), but this interpretation is challenged by recent findings showing within-class, and indeed within-language, discrimination on the basis of timing cues alone (White, Mattys & Wiget, 2012; see also Arvaniti & Rodríguez, 2013). Such results suggest that – rather than discrete categories – language discrimination relies on gradient variation along prosodic dimensions, with differences in speech rate and utterance-final lengthening shown to be particularly salient (adults: White et al., 2012; infants: White, Floccia, Goslin & Butler, 2014).

We used a head-turn preference paradigm to test whether English five-month-old infants could discriminate between French and Spanish. Although these have both been categorised as “syllable-timed”, they differ markedly in the realisation and distribution of strong syllables. We habituated infants to one of the two languages for at least 80 seconds and then exposed them to samples of new speakers, either from the same language or the other language. Infants reliably showed longer looking times to the new language compared with the familiar one, indicating that they perceived them as distinct (Figure 1).

Further experiments showed, however, that infant discrimination of even distantly-related languages is not straightforward: based on the same paradigm, English five-month-olds failed to discriminate either French/Finnish or Spanish/Finnish. Overall, this selective pattern of discrimination strongly argues that infant perception is not determined by intrinsic sensitivity to discrete classes. Rather, we argue that language discrimination reflects exploitation of functionally important – but gradient – differences in prosodic features such speech rate, final lengthening and lexical stress.
Figure 1: Mean orientation time for familiarised (same) and new languages.

References


Processing at different levels of the prosodic hierarchy: Dutch-learning and Turkish-learning infants
Brigitta Keij & René Kager
Utrecht Institute of Linguistics OTS, Utrecht University
b.m.keij@uu.nl, r.w.j.kager@uu.nl

In this study we compare infants learning rhythmically opposed language, namely Dutch-learning infants (with a dominant SW pattern similar to English) and Turkish-learning infants (with a dominant WS pattern similar to Hebrew). We use an innovative looking-while-listening eye-tracking procedure which presents infants with single pseudo-word (/nəldaf/ vs. /nəl’daf/) stimuli spoken by a speaker Spanish, in order to avoid an advantage for one of the language groups. Do Dutch- and Turkish-learning infants show a native rhythmic preference and at what age does this preference appear? In total, 90 Dutch- and 90 Turkish-learning infants aged 4, 6 and 8 months have been tested. The results of the Dutch-learning infants show a SW preference, which is strongest at 6 months of age (F (1,3443) = 7.170, p=.007). However, the Turkish-learning infants also show a SW preference, which is strongest at 4 months age (F (1,3443) = 11.992, p=.001).

We interpreted these results in light of a study investigating Hebrew-learning infants at 9 months of age (Segal & Kishon-Rabin, 2012). This study shows that Hebrew-learning 9-month-olds demonstrate a WS preference when listening to a speaker of Hebrew, while they showed a SW preference when listening to speaker of English, a foreign language. Consequently, it seems to matter whether infants are listening to a speaker of the native language or a foreign language. We proposed that infants resort to a ‘default’ SW preference when they do not recognize the speaker as a speaker of their native language. Alternatively, we proposed a native language dependency hypothesis. This hypothesis could explain the Hebrew results of as well as our own results. Therefore, adequate testing of language-specific rhythmic preferences probably requires a speaker of the native language.

In this paper we present the results from a follow-up study testing 48 Dutch- and 48 Turkish-learning infants at 6 and 8 months of age with a speaker of their native language. The results, however, are a replication of the earlier results: the Dutch-learning infants demonstrate a SW preference at 6 months of age and the Turkish-learning infants do not demonstrate a preference for either pattern at this age. These results suggest there is no influence of the language of the speaker. Consequently, the results from both studies should be interpreted differently. When we look at the level of the prosodic hierarchy on which the infants might have processed the stimuli, it seems plausible that pre-lexical infants at 4 months of age process the stimuli as phrases, while lexical infants from 6 months of age process the stimuli as words. On the phrase level, Turkish is actually trochaic, while Dutch is mixed due to word order variation. From this perspective, Dutch-learning infants not demonstrating a clear preference for either rhythmic pattern at 4 months of age (pre-lexical), while showing a SW preference at 6 months of age (lexical), and Turkish-learning infants presenting a SW preference at 4 months of age (pre-lexical), while not showing a clear preference for either rhythmic pattern at 6 months of age (lexical), makes sense.

Prosody perception and realisation in Mandarin Chinese: Data from children and adults
Hui-Ching Chen, Barbara Höhle
IDEALAB (International Doctorate for Experimental Approaches to Language and Brain)
University of Potsdam, Germany
huiching.chen@gmx.de

Over decades, there are numbers of studies investigating how information structure (IS) affects the structure of verbal utterances and how listeners interpret the cues related to IS [1][2][3]. But only recently, the question of how children acquire the linguistic means to mark IS has attracted more scientific interest. This paper deals with the acquisition of prosodic means of focus marking in a cross-linguistic perspective comparing German and Mandarin learning children.

German like English or Dutch belongs to the languages that mark focus mainly by prosodic means. However, previous research has shown that children - despite their evidenced sensitivity to prosodic information already present in infancy [4][5] - do not systematically use prosodic information for decoding IS in comprehension up to school age[6]. In Mandarin - as a tone language - the relevance of prosodic means to mark IS is more questionable. Previous research indicates that Mandarin speakers (adults and children) apply acoustic parameters i.e. fundamental frequency, duration and amplitude to mark IS in sentence production [6][7]. Most interestingly, Chen found that Mandarin learning children relied more heavily on stress than adults. She assumed that children switch their strategy from a prosodic one to a structural one based on word order across language acquisition [8].

These findings suggest a picture with an increasing ability to interpret prosodic markers of IS in children learning a language in which these markers are important while this ability decreases in children learning a language in which these markers are less important. In order to compare the developmental trajectory in producing and comprehending prosodic means to mark focus in children we tested German or Mandarin learning children and adults in a task that combined testing comprehension of prosodic markers of contrastive focus and its realization in production. In this task participants were presented with sentences that carried a pitch accent as a prosodic focus marker either on the sentence subject or the sentence object together with a picture. The pictures were set up in a way allowing for both a subject or object correction. We expected that the choice of subject or object correction would depend on the accentuation in the presented sentence with a focus accent on the subject leading to more subject corrections than a focus accent on the object. In each language 3, 4 and 5 year old children as well as a group of adults were tested.

We found in both languages a general preference to correct the object in children as well as adults. However, in German more subject corrections occurred when the subject carried a focus accent compared to the sentences in which the object carried the focus accent with no differences between the age groups. In contrast, for the Mandarin speaking participants no effect of prosody occurred: the amount of subject corrections was not affected by the accentuation of the test sentence across the age groups tested. This result correlates with the theoretical assumption that in Mandarin Chinese the most salient element is imbedded in the right most periphery of a sentence [9]. Second, even the youngest children (3-year-old) performed like Mandarin adults. That is MC speakers follow the structural cues instead of the prosodic cues from an early age on which is in contrast to the findings by Chen.
Cross-linguistically our findings suggest that children attune to the specific means of IS marking in their ambient language at a very early age.

**Target examples in the task:**

Simple NP condition
(1) Subject accented condition:
XIAONIAO yo pingzi, shi me?
'The BIRDY has the bottle, right?'
(2) Object accented condition:
Xiao-niao yo PINGZI, shi me?
'The birdy has THE BOTTLE, right?'

Complex NP condition
(3) Subject accented condition:
LUEXE DE SHANYANG yo xianggiao shi me?
'THE GREEN GOAT has the banana, right?'
(4) Object accented condition:
Luex de shanyang yo XIANGGIAO, shi me?
'The green goat has THE BANANA, right?'

**References**

Linking pitch prosody and affective behavior in mother-child interaction

Jennifer Cole¹, Daniel Berry², Nancy McElwain³, Timothy Mahrt¹, Nyssa Bulkes¹, Helen Emery³
¹Department of Linguistics, ²Department of Educational Psychology, ³Department of Human &
Community Development, University of Illinois
{jscole | djberry | mcelwn | tmahrt2 | bulkes2 | hemery2}@illinois.edu

In acquiring their native intonation system, a child must learn the form-function mapping, e.g.,
determining which pitch patterns convey what kinds of linguistic meaning. For pitch, this
relationship is arguably rooted in “biological codes” that reflect effects of speaker’s anatomy and
physiological state on speech production, and specifically, on F0 variation [1, 2]. These
biological codes give rise to universal patterns in the affective use of intonation (e.g., high pitch
conveys friendly affect, wide pitch excursions convey helpfulness), although the form-function
relationship is also subject to considerable refinement and specialization resulting in cross-
linguistic variation in the grammatical use of pitch. Here we test the hypothesis that the
biological code shapes the pitch prosody of mother’s child-directed speech, forging a direct link
between speech and affective behavior [3].

Method: Mothers (N = 62) and their 33-month-old children were observed during a well-
validated 15-minute laboratory play task. From audiovisual recordings, we assessed in 30-second
epochs: maternal and child positive affect, and maternal F0 (epoch-mean and –range F0).
Behavioral coding of positive affect was based on overt expressions of touch, kissing, smile,
laughter and enthusiastic verbal affirmation [4] but did not explicitly rely on maternal prosody.
Coding was judged reliable by inter-rater correlation (.88 (mothers), .85 (children)). Intervals of
mother’s speech were manually segmented and F0 was extracted using Praat [5], with pre- and
post-processing to adjust for mother’s F0 range and pitch-tracking errors.

Results: Cross-series correlations between behavioral measures of positive affect and F0
measures of maternal speech were positive and strongest within-epoch, indicating a closely-
timed relationship between pitch prosody and affective behavior. Linear mixed models tested:
(1) the extent to which temporal shifts in maternal behavioral affect were predictive of
contemporaneous changes in pitch prosody, and (2) whether these shifts in maternal pitch
prosody were uniquely predictive of shifts in child behavioral affect, after adjusting for the broad
effects of maternal expressed positive affect. Two main results were observed.

Maternal Expressed Positive Affect → Maternal Prosody: Controlling for speech
duration, speech rate and autoregressive processes in F0, within-person temporal increases in
maternal positive affect were associated with contemporaneous increases in mean F0 (B = .49, p
<.001; β = .10) and F0 range (B = 1.26, p < .001; β = .13).

Maternal Prosody → Child Expressed Positive Affect: There was also evidence of a
positive within-person relation between F0 and child affect. Temporal increases in F0 mean (B =
.013, p = .03; β = .08) or F0 range (B = .01, p = .02; β = .07) were associated with
contemporaneous increases in child positive affect. Although modest in absolute terms, these
respective F0 effects were unique after adjusting for the effects of observer-rated maternal
positive affect, speech duration, speech rate and autoregressive processes in child positive affect.
Collectively, our findings suggest that: (1) pitch prosody serves as important mechanism through
which mothers relay affective information to young children, and (2) maternal pitch prosody may
have direct effects on child affect that extend beyond broad observer-rated indices of maternal
affect.


Ambient language effects in the prosodic development of simultaneous bilinguals
Elaine Schmidt & Brechtje Post
University of Cambridge
aeis2@cam.ac.uk, bmbp2@cam.ac.uk

Previous studies analysing prosodic development in bilingual children suggest that the bilinguals’ languages are intermediate at early stages of development, situated somewhere between those of monolinguals of the two respective languages ([1]; [2]; [3]; [4]). It is only at the age of 4 that the languages become rhythmically separable ([5]; [3]; [4]). However, findings for the further developmental path are conflicting. While [1] observed characteristics of German uneven-timing in the Spanish of German-Spanish bilinguals, the bilinguals of [2] showed a preference for more Spanish even-timing in their English. The seemingly contradictory results might be the result of ambient language effects, since the 3-year-olds of [1] lived in Germany, whereas [2]’s participants were growing up in a Spanish speaking community in the US. If so, it shows the increasing importance of ambient language in rhythmic development from around the age of 3-4 in bilinguals. While 2-year-olds still show a very clear preference for even-timing, 3-year-olds seem to lose this bias and instead alter their speech to accommodate the properties of the ambient language. However, in order to tease apart ambient language effects properly, it is necessary to analyse and compare bilingual populations with the same experimental set up in both countries.

In this study we compared rhythm development in Spanish-English 4-, and 6-year-old bilinguals living in Spain (SPBL) and England (UKBL). We applied rhythm metrics (%V, nPVI-V, rPVI-C) to speech collected in semi-structured elicitation tasks in which children had to describe simple activities that were presented as animated images.

Results show that ambient language indeed has an effect on the simultaneous acquisition of two rhythmically different languages, but not in all rhythm metrics. We find a strong effect of ambient language in the variability of vocalic, but not consonantal material. Specifically, UKBL show too much variability in Spanish whereas SPBL show too little variability in English. English is characterised by high variability in vocalic durations as a result of lengthening and reduction of vowels depending on syllable type. Consequently, the high variability of UKBLs Spanish can be attributed to the rhythmic characteristics of English (Fig.1). In contrast, Spanish shows lengthening effects to a smaller extent and does not employ vowel reduction at all. The smaller variability found in SPBL speaking English (Fig.2) therefore clearly shows the influence of Spanish. However, unlike in English, the ambient language effects in Spanish disappear at in 6-year-olds.

We hypothesise that the reason for why we only find ambient language effects in vocalic metrics is that consonant production is more heavily dependent on fine-tuned motor control than vowel production ([6]; [7]; cf.[8]). Instead, the variability of vocalic intervals appears to be the result of the language-specific marking of accented and phrase-final syllables, where ambient language effects are also present (Fig.3).

References
Congress for the Study of Child Language and the Symposium on Research in Child Language Disorders. Madison.


Uncertainty can be encoded through various verbal and nonverbal means (lexicon, grammatical particles, prosody and gesture), which vary depending on the language. Studies focusing on the acquisition of the lexical encoding of the pragmatic meaning of uncertainty showed that children can distinguish between expressions of speaker certainty and uncertainty at age four (Moore et al., 1990; Matsui et al., 2003; Krahmer & Swerts, 2005). Furthermore, children’s ability to comprehend uncertainty has been shown to be correlated with their success in theory of mind tasks (i.e. the ability to attribute mental states, beliefs and desires to oneself and to others (e.g. Moore et al., 1990 / 1993; Matsui et al., 2009). The above-mentioned studies concentrated on the role of lexical cues in the understanding of uncertainty, while less is known about the role of prosody and gesture patterns. Studies on phonological development reported early infants' use of prosodic cues for pragmatic purposes before they are able to use lexical information (Esteve-Gibert & Prieto, 2013; Papaeliou & Trevathan, 2006; Snow, 2006). Furthermore, the role of facial gestures has been highlighted and shown to provide children with the scaffolding to linguistic meaning (Armstrong et al., 2014).

The aim of the present study is to investigate how gestural patterns interact with lexical and prosodic cues in the development of children’s ability to understand uncertainty. The following research questions are addressed: (1) Can children detect uncertainty more easily through gestures or the lexicon? (2) Can children detect uncertainty more easily through gestures or prosody? We hypothesize that gesture in interaction with prosody plays a more important role than the lexical content when children discover uncertainty.

A total of ninety 3- to 5- year-old Catalan-dominant children are tested in one of two experiments. Children watch a Powerpoint presentation in which two pairs of twins are guessing a third-character’s favourite fruit/sport/instrument. One of the twins knows the answer and the other one is uncertain, and the child has to guess the latter. In Experiment 1, the uncertainty meaning is expressed through lexical cues (modal verbs), accompanied by a facial expression of uncertainty and neutral falling contour (L*L%), whereas in Experiment 2, it is expressed through a rising intonation contour (L*H% or L+H*H%) accompanied by an uncertain facial expression. The Powerpoint contains the following trail types: audio-only, video-only and audio-visual (manipulated within subjects). Additionally, as a control task, a classical false-belief task is administered to the children (i.e. Sally-Ann task (Baron-Cohen et al., 1985) as well as an emotion detection task (i.e. Ruffman, 2002)).

Preliminary results from 47 children performing either Experiment 1 or 2 suggest that contrary to previous results, children start to recognise uncertainty already at age three. Comparing results of the two experiments, it can be seen that children perform better when they have prosodic and gestural cues at hand (Experiment 2). It seems that they provide children with the scaffolding for linguistic meaning of uncertainty.
References


Can 6-year-olds use prosodic cues to disambiguate compounds from lists?

Yuen, I., Xu Rattansone, N., Schmidt, E., Demuth, K.
Macquarie University

ivan.yuen@mq.edu.au, nan.xu@mq.edu.au, elaine.schmidt@mq.edu.au, katherine.demuth@mq.edu.au

Although young children are sensitive to the rhythmic and prosodic patterns of their native language, they do not appear to employ these cues in differentiating compounds from phrases in production [1]. According to [9], the difficulty might be related to children’s use of prosodic information in perception. However, children as young as 2 are reported to lengthen syllables at the ends of words and phrases [8, 10], suggesting that they can use duration to indicate prosodic structure. This then raises questions about what prosodic cues children might use to distinguish compounds from phrases in production, and if these are employed in an adult-like fashion.

Adults use both duration and pitch to disambiguate Noun (N1) – Noun (N2) compounds vs. lists [4, 5], where both N1 and N2 undergo polysyllabic shortening in the compound [3], but not in the list. This leads to a shorter duration of N1 and N2 in the compound. As only N2 undergoes word-boundary lengthening in the compound, the durational difference between N1 and N2 within the compound provides additional cues to structure [6] [7]. At the pitch level, both N1 and N2 tend to exhibit a continuation rise boundary in a list, however, only N2 is likely to do so in the compound. We therefore predicted a higher mean f0 on N1 in the list than in the compound. On the basis of [8, 10], we then expected children to exhibit boundary-related lengthening on N1 and N2 in the list, but only on N2 in the compounds. Just as in adults, we also predicted children would show a shorter N1 in the compound than in the list. It is also possible that children might show polysyllabic shortening in the compound. Given [9], we did not expect the children to employ disambiguating pitch cues in their productions.

Eight Australian English-speaking children (7F, 1M, mean age = 5.8 years) and adults (6F, 2M, mean age = 18.6 years) participated in an elicited production task. The target nouns contained seven compounds and lists (e.g., ice-cream and X vs. ice, cream, and X). An Australian English-speaking experimenter showed the stimuli in pictures and asked the participants what they saw. They were instructed to respond, using the carrier sentence ‘I can see …’. All productions were audio recorded and acoustically analysed using Praat [2].

As predicted, both children and adults showed a significant difference in N1 and N2 durations, suggesting a boundary-related lengthening on N2. The results also suggest both populations are using polysyllabic shortening, as both N1 and N2 were significantly shorter in the compounds compared to the lists (Fig.1). Interestingly, the 6-year-olds also showed a significant difference in mean f0 on N1 between the compounds and lists (Fig. 2), like the adults. This suggests that at least some 6-year-olds are using adult-like prosodic cues to distinguish compounds from lists.
References


---

Fig. 1.: Mean duration of nouns in compounds and lists in children and adults

Fig. 2. Mean f0 of nouns in compounds and lists in children and adults