Published in final edited form as:

J Mem Lang. 2015 July 1; 82: 105–117. doi:10.1016/j.jml.2015.03.005.

The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns

Kara Hawthorne^a, Reiko Mazuka^b, and LouAnn Gerken^c

^aUniversity of Arizona, Department of Linguistics 1100 E University Blvd, Tucson, AZ 85721;
University of Alberta, Department of Linguistics 2-40 Assiniboia Hall, Edmonton, AB, Canada T6G 2E7; khawthor@ualberta.ca

^bRIKEN Brain Science Institute, Information Science Building 2-1 Hirosawa Wako City, Saitama 351-0199, Japan; Department of Psychology and Neuroscience Box 90086, 417 Chapel Drive, Duke University, Durham, NC 27708-0086 mazuka@brain.riken.jp

^cUniversity of Arizona, Department of Psychology 1503 E University Blvd, Tucson, AZ 85721; gerken@email.arizona.edu

Abstract

There is mounting evidence that prosody facilitates grouping the speech stream into syntactically-relevant units (e.g., Hawthorne & Gerken, 2014; Soderstrom, Kemler Nelson, & Jusczyk, 2005). We ask whether prosody's role in syntax acquisition relates to its general acoustic salience or to the learner's acquired knowledge of correlations between prosody and syntax in her native language. English- and Japanese-acquiring 19-month-olds listened to sentences from an artificial grammar with non-native prosody (Japanese or English, respectively), then were tested on their ability to recognize prosodically-marked constituents when the constituents had moved to a new position in the sentence. Both groups were able to use non-native prosody to parse speech into cohesive, reorderable, syntactic constituent-like units. Comparison with Hawthorne & Gerken (2014), in which English-acquiring infants were tested on sentences with English prosody, suggests that 19-month-olds are equally adept at using native and non-native prosody for at least some types of learning tasks and, therefore, that prosody is useful in early syntactic segmentation because of its acoustic salience.

Keywords

prosody; prosodic bootstrapping; constituents; syntax acquisition; language acquisition; Japanese; English

© 2015 Published by Elsevier Inc.

Correspondence concerning this article should be addressed to Kara Hawthorne, Department of Linguistics, University of Alberta, 2-40 Assiniboia Hall; Edmonton, AB T6G 2E7; Canada. E-mail: khawthor@ualberta.ca. Telephone: (780) 492-5174. Fax: (780) 492-3434..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The prosodic bootstrapping hypothesis (Gleitman & Wanner, 1982; Morgan, 1986; Peters, 1983) argues that prosody – speech melody and rhythm – may be useful to infants as they first endeavor to parse a continuous stream of phonetic information into words and syntactic constituents. Indeed, stress and other prosodic cues facilitate early word segmentation in English (Gout, Christophe, & Morgan, 2004; Jusczyk, Cutler, & Redanz, 1993; Shukla, White, & Aslin, 2011). Likewise, the prosodic correlates of syntactic boundaries aid early syntactic segmentation (e.g., Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000). Prosody continues to impact processing in 20-month-olds (Soderstrom & Morgan, 2007), as well as word segmentation and syntax processing in older children (Beach, Katz, & Skowronski, 1996; Snedeker & Yuan, 2008) and adults (Christophe, Gout, Peppercamp, & Morgan, 2003; Langus, Marchetto, Bion, & Nespor, 2012).

In this paper, we focus on prosody's role in helping learners segment speech into syntactic constituent candidates. The boundaries of major syntactic constituents, such as clauses, correlate with the prosodic cues of pauses, final-syllable lengthening, and pitch resets. These cues are particularly salient and reliable in infant-directed speech (e.g., Soderstrom, Blossom, Foygel, & Morgan, 2008). Recent work suggests that prosody supports syntactic segmentation in infants and may even underlie infants' early conceptualizations of syntactic constituency (Hawthorne & Gerken, 2014; Nazzi et al., 2000; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003; Soderstrom, Kemler Nelson, & Jusczyk, 2005). However, the mechanism that supports prosody's potential role in syntax acquisition and early language processing is unknown. What is perceived by an adult native speaker as prosodic information about stress and intonation may be interpreted by a naive listener merely as acoustic information about the relative pitch, duration, and amplitude of sounds. These acoustic cues are perceivable by almost anyone possessing a mammalian auditory system, but the way they interact to manifest as a syntactic grouping, pitch accent, or contrastive stress varies based on the language in question and therefore must be learned.

We consider two hypotheses about the mechanism underlying the prosodic bootstrapping of syntactic constituents. First, the *acoustic salience* hypothesis asserts that prosody is useful to language learners simply because it is acoustically salient. Human newborns (Nazzi, Bertoncini, & Mehler, 1998), cotton-top tamarins (Ramus, Hauser, Miller, Morris, & Mehler, 2000), and rats (Toro, Trobalon, Sebastian-Galles, 2003) can all discriminate between languages that differ along the prosodic dimension of rhythm. This suggests that the cues underlying perception of linguistic rhythm are easy for mammals to perceive without any prior language-specific experience – an important prerequisite for a bootstrapping candidate.

Further supporting this hypothesis, French newborns who were habituated to a disyllabic sequence extracted word-medially or straddling a word boundary dishabituated to the disyllable at test when the prosody switched from word-internal to boundary-straddling. They did so both when the disyllable was extracted from a French carrier sentence, and when it was extracted from a Spanish carrier sentence (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, & Sebastian-Galles, 2001). This supports the idea that

native and non-native prosody influence spoken language processing similarly, at least for very young listeners. If prosody's general acoustic manifestations are sufficient to support learning, experience with the target language should not be required for listeners to use prosody to parse speech into syntactic constituent-like chunks. Indeed, Langus and colleagues (2012) found that adult Italian speakers could use synthesized Italian-like and Japanese-like prosody to learn the structure of a hierarchically-organized artificial grammar.

On the other hand, it is also true that adults interpret prosody as linguistically meaningful in ways that vary from language to language. For example, English speakers interpret sentence-final rising pitch as question intonation. This knowledge is likely not present at birth, since speakers of some languages use flat or even falling intonation for questions (e.g., Iivonen, 1998). In support of this hypothesis is evidence that native-language prosody impacts speech perception within the first few months of life – certainly before infants are likely to be learning much about syntax. While newborns cannot discriminate between their native language and another language in the same rhythmic class, five months later they can (Nazzi & Remus, 2003). French and German 4-month-olds show a processing advantage for words with native-like stress patterns (Friederici, Friedrich, & Christophe, 2007), and native phrasal prosody even influences the prosodic characteristics of 18-month-old French and Japanese infants' disyllabic babbles (Halle, de Boysson-Bardies, & Vihman, 1991). If it is the case that the mechanism underlying prosodic bootstrapping relies on the listener's acquired knowledge of her native language, experience with the target language would be an important prerequisite to using prosody to parse speech into syntactic constituents. This is the *language-specific* hypothesis.

Evidence for the language-specific hypothesis comes from studies comparing infants' performance with native and non-native prosody in tasks that tap the prerequisites of syntax learning. For example, English-acquiring infants can perceive the prosodic correlates of major syntactic boundaries in their native language (Hirsh-Pasek et al., 1987; Jusczyk, 1989; Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989), but evidence is mixed for infants who are listening an unfamiliar language. Jusczyk (1989) found that English-acquiring 4.5-month-olds, but not 6-month-olds, preferred German speech when pauses occurred at clause boundaries rather than inside the clauses, while Fernald and McRoberts (1996) found no evidence of a similar preference for English-acquiring 4-, 7-, or 10-month-olds listening to Polish. This suggests that infants listening to an unfamiliar language may have trouble detecting violations of the typical correspondence between pauses, final lengthening, and pitch resets at clause boundaries.

Similarly, English-acquiring infants who are listening to their native language are better at remembering strings of words that form prosodic constituents rather than the same string of words spoken with list prosody or straddling a prosodic boundary (Mandel, Jusczyk, & Kemler Nelson, 1994; Mandel, Kemler Nelson, & Jusczyk, 1996; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Seidl, 2007; Soderstrom, Kemler Nelson, & Jusczyk, 2005; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Johnson and Seidl (2008; Seidl, 2007) found that both English- and Dutch-acquiring 6-month-olds recognized a familiarized prosodically-signaled clause at test when listening to their native language, but they also mention an unpublished experiment¹ in which neither English-acquiring infants listening to

Dutch nor Dutch-acquiring infants listening to English were able to recognize a non-native clause at test. Thus, despite the salience of prosodic information, non-native clausal prosody may not facilitate chunking the speech stream in the same way as native prosody for 6-month-olds, even across prosodically-similar languages. However, it would be unwise to draw strong conclusions from these two null results.

In this paper, we test the acoustic salience and language-specific hypotheses through two experiments that probe whether 19-month-old infants are equally adept at parsing an artificial grammar into syntactic constituent-like units when listening to native versus non-native prosody. These experiments build on Experiment 1 from Hawthorne and Gerken (2014, henceforth referred to as H&G). In H&G, American English-acquiring 19-month-olds listened to sentences from an artificial grammar with 1-clause (ABCDEF) or 2-clause (ABC, DEF) English prosody. Both groups were then tested on sentences that represented a grammatical (DEF, ABC) or ungrammatical (EFA, BCD) reordering or ‘movement’ of the prosodically-marked ‘clauses’ from the 2-clause familiarization. Infants in the 2-clause group discriminated between the grammatical- and ungrammatical-movement items, showing a listening preference for sentences that preserved the clauses heard at familiarization. Infants in the 1-clause (control) group did not discriminate between the test item types. This suggests that infants can use their native prosody to locate groups of words that they treat like cohesive, moveable units that are independent from a particular acoustic manifestation – much like syntactic constituents.

In the current paper, we ask whether non-native prosody similarly facilitates parsing speech into constituent-like units. In Experiment 1, we test Japanese-acquiring infants on the same English-based stimuli used in H&G; in Experiment 2, we test English-acquiring infants on Japanese-based stimuli. Coupled with results from H&G, these experiments address the role of ambient language experience in shaping the prosodic bootstrapping of syntax. Japanese and English are prosodically distinct on a number of dimensions, so if learners can use unfamiliar prosody to bootstrap knowledge of constituent-like groupings, our results would support the acoustic salience version of the prosodic bootstrapping hypothesis and suggest that the acoustics of prosody are sufficient to facilitate syntactic learning. On the other hand, if learners cannot use non-native prosody to locate syntactic constituents, our results would support the language-specific version of the prosodic bootstrapping hypothesis and suggest that prior experience with the correlations between prosodic cues and syntactic boundaries in the target language is a necessary prerequisite for the prosodic bootstrapping of constituent-like units.

Nineteen-month-olds were selected for these experiments because infants at this age have significant knowledge of their native language, including its prosody. In fact, 18-month-olds even use language-specific phrasal prosody in their disyllabic babbles (Halle et al., 1991). If the language-specific hypothesis is true, then 19-month-olds should be sufficiently familiar with the prosodic patterns of their language that they would be confused by a new prosodic system. To mimic the situation faced by younger infants, the auditory stimuli in the present experiments were from an artificial grammar with novel vocabulary and syntactic structure,

¹Confirmed through personal communication with E. Johnson & A. Seidl, March 10, 2015.

as well as non-native prosody. The 19-month-olds tested here are in a situation that is perhaps akin to that of an internationally-adopted child who is exposed to a new language after early infancy.

Before describing the experiments, a brief comparison of English and Japanese prosody is in order.

1.1 A Comparison of English and Japanese Prosody

Clauses typically correspond to intonational phrases (IPs; see Nespor & Vogel, 2007), which are marked with pauses, lengthening of preboundary syllables, and pitch resets. IPs share certain similarities across languages, likely due in part to the physiology of breathing. Indeed, both English (Broen, 1972) and Japanese (Kawahara & Shinya, 2008) speakers are more likely to pause in intonational phrase (IP)/clause-final position than elsewhere in the sentence (c.f., Cruttenden, 1986). American English and Japanese parents also both exaggerate pauses when using infant-directed speech (IDS) (Fernald et al., 1989).

English and Japanese each exhibit final-syllable lengthening, though to different extents. English has a relatively strong degree of final lengthening, particularly in IDS (Cooper & Paccia-Cooper, 1980; Soderstrom et al., 2008). It has been argued in the past that Japanese does not have final lengthening (e.g., Nishinuma, 1979; Takeda, Sagisaka, & Kuwabara, 1989; Venditti & Tamashita, 1994), but more recent evidence suggests that final lengthening does occur (Martin, Igarashi, Jincho, & Mazuka, under review; Ueyama, 1999). However, there is evidence that the degree of final lengthening in Japanese is less than is seen in English. For example, one study of spontaneous IDS found that American English-speaking mothers produced utterance-final vowels an average of 2.41 times longer than non-utterance-final vowels, compared to a ratio of 1.75 to 1 for Japanese mothers (Fisher & Tokura, 1996). Even at 18 months, Japanese infants show less final lengthening in their productions than French infants (French, like English, has relatively extreme final lengthening) (Halle, de Boysson-Bardies, & Vihman, 1991).

Pitch also operates quite differently in English and Japanese. The languages have a similar degree of pitch excursions in pre-pausal positions (Fisher & Tokura, 1996), but differ in the types of intonational contours that occur and in how pitch is used. In English, pitch accents can change meaning at the utterance level (“John is running,” versus “John is RUNning [not walking]”; pitch accent is marked in all caps), and the pitch contour of an utterance can be manipulated to express things like uncertainty and impatience (Pierrehumbert, 1980). In Japanese, an inventory of phrase-final Boundary Pitch Movements (BPMs) expresses various pragmatic distinctions (e.g., a rising boundary tone is typically used to signal a question) (Beckman & Pierrehumbert, 1986; Igarashi, Nishikawa, Tanaka, & Mazuka, 2013; Venditti et al., 2008). BPMs aside, pitch contours of Japanese utterances are less variable than those in English, since they are largely determined by lexically determined high-falling pitch accents (Igarashi et al., 2013). Also, Japanese IDS has been reported to involve less pitch exaggeration than American English IDS (Fernald et al., 1989), though recent evidence suggests that speakers of both languages do exaggerate pitch when speaking to infants: American English speakers exaggerate pitch accents, while Japanese speakers expand the pitch range of BPMs (Igarashi et al., 2013).

Finally, English and Japanese differ rhythmically. English is considered a stress-timed language and Japanese is a mora-timed language. The rhythmic differences between the two languages also extend to the level of syntactic constituents. Indeed, Nespor et al. (2008) argue that English and Japanese have different templates for phrasal prosody because of the iambic/trochaic law and the complement law. The iambic/trochaic law is a perceptual phenomenon whereby elements of alternating duration are perceived as having the most prominent (longer) element last, while elements of alternating intensity are perceived as having the more prominent (louder) element first. The complement law is the tendency for the complement to be more perceptually prominent than its head (Nespor & Vogel, 2007). Verbs precede objects (their compliments) in English, so the ends of English predicates tend to be marked with duration cues. This has been confirmed by studies finding relatively strong final-syllable lengthening in English. In Japanese, objects precede verbs, so the beginnings of Japanese predicates tend to be louder, with a subsequent decrease in amplitude (c.f., Fisher & Tokura, 1996). The dominant phrase-level duration patterns in the ambient language is suggested to affect infant perception, even at the fundamental level of grouping biases. At 5-6 months neither English- nor Japanese-acquiring infants segment sequences of alternating short and long tones into either iambic (short-long) or trochaic (long-short) chunks; by 7-8 months, English-acquiring infants (like English-speaking adults) prefer iambic groupings, while Japanese-acquiring infants (like Japanese adults) still show no consistent preference (Yoshida et al., 2010).

There are, therefore, numerous prosodic differences between English and Japanese that affect the manifestation of prosodic cues to major syntactic boundaries. If prosodic bootstrapping relies on specific acquired knowledge of the prosody-syntax interface in the target language, English and Japanese listeners should fail to learn from the other language's prosody. Despite differences, though, both languages share core acoustic cues – pauses, some degree of final lengthening, and pitch resets at clause boundaries. If prosodic bootstrapping is a general acoustic phenomenon based solely on salience, there are sufficient boundary cues in both English and Japanese to facilitate learning.

2. Experiment 1

In Experiment 1, we familiarized Japanese 19-month-olds with the same English-based nonce-word sentences used in Experiment 1 of H&G. Half of the infants were familiarized with sentences with 2-clause prosody (ABC, DEF) and half with 1-clause prosody (ABCDEF). During the test period, infants heard sentences that represented a grammatical (DEF, ABC) or ungrammatical (EFA, BCD) movement of the clauses from the 2-clause familiarization.

If the acoustic salience hypothesis is correct, the perceptual prominence of pausing, final-syllable lengthening, and pitch resets in the English-based sentences may be sufficient to induce learning in Japanese infants. If this is the case, Japanese infants should behave like their American counterparts in H&G, with those in the 2-clause group discriminating between the grammatical- and ungrammatical-movement test items.

As in H&G, the 1-clause group serves as a control for the possibility that learners preferentially attend to boundary words due to primacy and/or recency effects. The 1-clause familiarization sentences share two boundary words with the grammatical-movement test items and zero with the ungrammatical-movement items, while the 2-clause familiarization sentences share four boundary words with the grammatical-movement items and none with the ungrammatical-movement items. If the infants base their listening preferences on recognition of boundary words, both the 1-clause and 2-clause groups should discriminate between the test item types, with the 2-clause group perhaps discriminating to a larger extent (see Section 2.4 and footnote 4, as well as H&G for additional discussion of this issue).

Finally, it is also possible that neither the 2-clause nor the 1-clause infants will discriminate between the test item types. English and Japanese have numerous differences in clause-level prosody, so the English-based stimuli will be unfamiliar to the Japanese infants. This result would support the language-specific hypothesis by suggesting that acquired knowledge about the target language's prosody is necessary to learn the experimental grammar.

2.1 Methods

2.1.1 Participants—Twenty-four monolingual Japanese-acquiring 19-month-olds living in Tokyo, Japan were tested (18 male, 6 female). Participants were all of normal birth weight (5lb, 7oz) and term (37 weeks). Data from an additional 19 infants were either not collected or excluded due to equipment error ($n = 1$), extreme shyness and crying during the long familiarization phase ($n = 3$), fussiness or crying during the 2-minute familiarization or testing phases ($n = 14$), and for standing with head out of camera range during testing ($n = 1$).

2.1.2 Materials—Materials were identical to those used with English-acquiring infants in H&G, Experiment 1. The sentences were recorded by a native English-speaking female from the midwestern United States, who was instructed to speak in an infant-directed register. Recorded sentences were partially standardized for amplitude, tempo, and pause length in order to avoid any potentially distracting difference between individual stimuli sentences. This was done by adjusting any value that was more than one standard deviation from the mean to the mean value for that measure.

Each familiarization sentence comprised six words with 1-clause (ABCDEF, modeled on “I ate at home with John”) or 2-clause (ABC, DEF, modeled on “While I ate, John came home”) prosody. The word classes always occurred in the same order (A-B-C-D-E-F), with each letter representing a class of two nonsense CVC word tokens (Table 1, below), for a total of 64 unique sentences for each familiarization prosody. Sentences containing particular ABC and DEF combinations (A1-B1-C1, A2-B2-C2, and A1-B2-A2; D1-E1-F1, D2-E2-F2, and D1-E2-F1) were withheld from familiarization to be used at test, to ensure that the infants were generalizing at the level of the word classes. Once those combinations were withheld, there were 25 unique familiarization sentences remaining. These were individually recorded and then placed into pseudorandomized familiarization lists (1-clause and 2-clause), with the sentences separated by a 1-second pause.

Since the nonce words were based on English, they violated Japanese phonology in several ways. First, all syllables were CVC, and a variety of consonants occurred in coda position. In Japanese, coda consonants can only be nasals or geminates with the onset of the next syllable in a multi-syllabic word. Second, the phonetic inventory used in the stimuli words included several phonemes that do not occur in Japanese. The consonants /f/ and /v/ only occur rarely and exclusively in loanwords; / / and /l/ do not occur at all. Phonemically, Japanese only has the tense vowels (/i, e, , o, a/), so the lax vowels used in most of the stimuli words (/ , , æ, / and the tense vowel / /) were not canonical native vowels. However, since the tense/lax dimension is not distinctive in Japanese, Japanese speakers exercise a certain amount of flexibility in their pronunciation, so the vowel sounds in the stimuli words should not be unfamiliar to the infants.

The familiarization sentences were acoustically analyzed for clause-level prosodic features. The mean intra-clause pause duration for the 2-clause items was 173ms, and there were no pauses in the 1-clause items. Clause-final words (C and F for the 2-clause items and F for the 1-clause items) had significant lengthening of the syllable nucleus compared to the non-clause-final words ($F(1, 57) = 38.06, p < .001$) (see Table 2 for descriptive results). Both the 1-clause and 2-clause sentences had an appropriate intonational contour for the intended prosody (see Figure 1). Note that the speaker placed a pitch accent on word D of the 1-clause items.

All test items had 2-clause prosody. Grammatical-movement test items were of the form DEF, ABC, and the ungrammatical-movement items were of the form EFA, BCD. Test sentences were each individually recorded, in order to ensure that the grammatical- and ungrammatical-movement test sentences were equally natural sounding. The clauses in the grammatical-movement items were grammatically the same as those that the 2-clause group heard during familiarization (ABC and DEF), though the clauses appeared in the opposite order. Because the clauses had ‘moved,’ they also had a new prosodic contour (e.g., ABC occurred with initial-clause prosody during 2-clause familiarization and with final-clause prosody for the grammatical-movement test items). The ungrammatical-movement items chunked the sentences into new, non-constituent units. Since the clauses used at test were withheld from familiarization, the particular word combinations were novel at test.

2.1.3 Design and task—Half of the infants were familiarized with 1-clause prosody and half with 2-clause prosody. Familiarization was divided into two phases. The initial, long familiarization occurred in a small room, where the infant, a parent, and an experimenter played quietly with toys while stimuli played in the background. This phase lasted ten minutes, after which the infant and parent moved into the testing booth, where they heard an additional two minutes of familiarization. While the total familiarization duration was the same for both groups, the 1-clause group heard 31 more familiarization sentences than the 2-clause group (261 versus 230) because the 1-clause sentences were shorter in duration. Since the 1-clause group is not predicted to discriminate between the test item types under either the acoustic salience or language-specific hypothesis, this works against our predictions.

During the 2-minute familiarization and testing phases, infants were seated on their parents’ lap in a soundproof booth. The parent wore headphones and was instructed to remain neutral

so as to not bias the infants' responses. Testing occurred using a modified version of the Head-Turn Preference Procedure (Kemler Nelson et al., 1995). There was a green light in front of the infant and two orange lights directly below speakers, one on each side wall. This is the same set-up used in H&G, except in that study the infant was seated in a highchair, with the parent slightly behind and to the side. This change was made because the Japanese infants were too fussy when seated in a highchair.

The entire 2-minute familiarization list played from both speakers simultaneously (cf. Saffran, Aslin, & Newport, 1996), while the light flashed according to the infant's direction of looking in order to allow the participant to learn the contingency between the lights and their looking behavior. An experimenter in an adjacent room, who could not hear the stimuli, monitored the infant's head turns. During the test phase, the infant heard 12 test trials, divided into three blocks, each with two grammatical-movement and two ungrammatical-movement test trials. The order of the test trials was randomized within each block. A test trial began when the infant oriented to the center flashing light. A side light would then begin to flash, and a test list began playing from the adjacent speaker once the infant turned his or her head to look at the light. The trial ended when the infant looked away for two seconds.

2.2 Results

Looking times were recorded for each trial, and, following standard practice in our lab, those that were shorter than two seconds were excluded from analysis to ensure that we did not include trials where the infants' attention was never captured, but she just turned her head toward the flashing light in passing. The mean looking time was 7.43s. Eight of the 12 infants in the 2-clause condition and five of the 12 in the 1-clause condition listened longer to the ungrammatical-movement test items.

Data were analyzed with linear mixed effects modeling. This type of modeling was chosen because it is able to deal with missing data points and can account for random effects of overall looking time as well as experimental manipulations. Data were analyzed using R (R Core Team, 2014), and the R packages *lme4* (Bates, Maechler, Bolker, & Walker, 2014) and *languageR* (Baayen, 2014). Familiarization condition (1- versus 2-clause) was a between-participants variable and test condition (grammatical- versus ungrammatical-movement) was a within-participants variable. Test condition was included as a random effect grouped by participants (slope), and participant was included as a random intercept. Looking time was the dependent variable, and looking times were log transformed to increase the normality of the residuals. Significance of the fixed effects was assessed using likelihood ratio tests comparing a full model to a minimally reduced model, so results are reported using the chi-squared (χ^2) statistic.

There was a significant interaction of familiarization by test condition ($\chi^2(1) = 8.65, p = .0033$). The effect of test condition was significant for the 2-clause condition ($\chi^2(1) = 4.66, p = .031$), with infants listening longer to the ungrammatical-movement test items. The effect of test condition was not significant for the infants in the 1-clause condition, although there was a non-significant trend towards a preference for the grammatical-movement test items ($\chi^2(1) = 3.22, p = .073$). Results are presented in Figure 2.

Closer examination of the data reveals that the trend towards a familiarity effect seen in the 1-clause group was largely caused by an odd pattern of looking times for the first test trial. This is reflected in the significant interaction between test condition and test trial type (first or non-first) seen in the 1-clause group ($\chi^2(2) = 10.15, p = .0063$). Infants in the 1-clause group whose first test trial was a grammatical-movement item listened much longer than average to that first test trial: a mean of 17.00s, which is more than 2.5 standard deviations above the overall mean looking time of 7.35s and over twice as long as the mean looking time for the first test trial (8.20s). Conversely, the infants in the 1-clause group whose first test trial was an ungrammatical-movement item had unusually short looking times – an average of 4.87s.²

To check the validity of the results without the potentially biasing influence of the first test trial, the data were reanalyzed with the first test trial omitted for all participants. As was the case when the first test trial was included, there was a significant interaction between familiarization and test condition ($\chi^2(1) = 5.48, p = .019$), and a significant effect of test condition for participants in the 2-clause group ($\chi^2(1) = 4.54, p = .033$). However, there is no longer a trend towards an effect of test condition in the 1-clause condition ($\chi^2(1) = 1.15, p = .28$).³

2.3 Discussion

As with Experiment 1 of H&G, there are a few ways to interpret the learning that the infants demonstrated here. Conservatively, our results suggest that infants treat prosodically-grouped words as more cohesive than words that straddle a prosodic boundary. Cohesiveness is an identifying property of syntactic constituents, so it is also possible that the infants interpreted the cohesive chunks as constituents. This interpretation is further supported by the fact that the infants recognized the clauses at test, even though the clauses (1) appeared in a transformed version of the original sentence, (2) had new acoustic contours (e.g., clause ABC had initial-clause prosody during familiarization and final-clause prosody at test), and (3) had new word combinations (the tokens of ABC and DEF used at test were not heard during familiarization).

Regardless of how much syntax the infants were inferring from the familiarization items, the current results support the acoustic salience version of the prosodic bootstrapping hypothesis. Experience with the target language is not necessary for prosody to be an effective cue to cohesive, constituent-like groupings. Despite the fact that the English-based

²It is difficult to interpret this first trial effect. The 1-clause participants experienced a change from 1-clause to 2-clause prosody at test, which could account for unusually long looking times; however, this interpretation is challenged by the short looking times from the 1-clause infants whose first test trial was an ungrammatical-movement item, since they also went from 1-clause prosody at familiarization to 2-clause prosody at test. Another possibility is that the infants from the 1-clause group with a grammatical-movement first test trial recognized that the clause-initial words from familiarization were still clause-initial in the consistent test items and were therefore showing a true familiarity effect. This would suggest that the infants in the 1-clause group would show an overall preference for the grammatical-movement test items, however, which is clearly not the case (see Figure 2). A third, more likely, possibility is that the effect is due to random variation. In order to be certain that the effect is due to noise, however, we would have to test more infants. Given that the 1-clause group did not show a statistically-significant familiarity effect overall, it would be unwise to conclude from these four extreme data points that the 1-clause group truly had a familiarity preference.

³The same pattern of results occurs if the data are analyzed with ANOVAs. There is a significant familiarization by test condition interaction ($F(1,22) = 11.93, p = .0023$), with a significant main effect of test condition for the 2-clause group ($F(1,11) = 5.70, p = .036$), but not the 1-clause group ($F(1,11) = 1.56, p = .22$). Note that there is no marginal effect for the 1-clause group, so we did not conduct an ANOVA analysis without the first item.

stimuli are phonotactically and phonemically very different from the Japanese participants' ambient language and the prosody of the stimuli – the pitch contours, in particular – is unfamiliar to the Japanese infants, participants in the 2-clause group were able to discriminate between the grammatical- and ungrammatical-movement items. Infants in the 1-clause group did not significantly discriminate between the test item types, suggesting that the 2-clause group did not base their looking time preferences solely on boundary words. Instead, it suggests that prosodic cues are sufficiently acoustically robust that they can be used to facilitate the acquisition of syntactic constituents or the underpinnings thereof without prior knowledge of the interactions between prosody and syntax in the target language.

The possibility that our results are based primarily on recognition of boundary words was more strongly tested in H&G, Experiment 2, in which the English-based artificial grammar described above was tweaked such that the familiarization items had 2-clause, 1-IP prosody, unlike the 2-clause, 2-IP prosody used in the present experiment and in H&G, Experiment 1. While the 2-clause, 2-IP prosody used here was modeled on “While I ate, John came home,” the 2-clause, 1-IP prosody of H&G, Experiment 2 was modeled on “John came home while I ate,” which does not have a pause between the two clauses. Like the 1-clause items, then, the familiarization stimuli for H&G, Experiment 2 shared two pre-/post-pausal words with the grammatical-movement test items (A and F). This is in contrast to the four shared pre-/post-pausal words between the 2-clause familiarization items used in the current experiment and the grammatical-movement test items (A, C, D, and F). Nonetheless, the 19-month-old English-acquiring participants in H&G, Experiment 2 successfully discriminated between grammatical-movement and ungrammatical-movement test items (identical to the ones used here, i.e., with 2-clause, 2-IP prosody), suggesting that infants are not primarily relying on recognition of words that occur in perceptually salient pre- and post-pausal positions. Note that this result also suggests that the failure of the 1-clause group is not because they were distracted by the change in prosody from familiarization to test, since the infants in H&G, Experiment 2 likewise experienced a change from 1-IP to 2-IP prosody.

An interesting pattern of results emerges when comparing the performance of the Japanese infants with the performance of the American infants in H&G, Experiment 1. In both cases, only the infants in the 2-clause group discriminated, but the direction of preference was different, depending on the infant's native language. The English-acquiring infants had a familiarity preference for items that had a grammatical movement of the clauses they heard during familiarization, while the Japanese-acquiring infants preferred the novel, ungrammatical-movement test items. Because the two groups of infants were tested in different labs by different research teams, it is risky to interpret the difference in listening preference. We did not *a priori* predict either a familiarity or novelty effect for these experiments. Indeed, it is the significant difference in listening times for the 2-clause infants and lack of significant difference for the 1-clause infants that are crucial for the current purposes.

We are tempted, however, to speculate on possible reasons for the difference between American and Japanese infants. One intriguing explanation is that the discrepancy in direction of preference is due to cultural differences (c.f., Fernald et al., 1989), which may

have mediated attention during the initial familiarization phase. It has been noted that Japanese mothers spend more time rocking and carrying their infants compared to American mothers (Caudill & Weinstein, 1969), and this was observed by the experimenters during the initial long familiarization phase. In H&G, every American infant played with toys during the entire ten minute period, while the majority of Japanese infants spent a portion – or all – of that time on their mothers' laps, watching the experimenter play with the toys. Since the Japanese infants were, on average, less actively engaged with the toys during this period, they may have spent more time listening to the familiarization stimuli, such that by the testing period they preferred to listen to the more novel, ungrammatical-movement test sentences.

Experiment 2, in which English-acquiring infants are tested on stimuli with Japanese-like prosody and phonology, will help solidify the interpretation of the direction of preference change between H&G and Experiment 1 of the present paper. If American infants in the 2-clause group of Experiment 2 continue to show a familiarity preference like the American infants in H&G, even while tested on non-native input, like the Japanese infants in Experiment 1, a cultural or attentional difference between the two groups of infants will remain a viable explanation.

More importantly, testing English-acquiring American infants on Japanese prosody will allow us to examine prosody-based learning of constituent-like chunks in a language that is very different from English.

3. Experiment 2

In Experiment 2, American English-acquiring infants were familiarized and tested on sentences with Japanese-like prosody and phonemes.

3.1 Methods

3.1.1 Participants—Twenty-four monolingual English-acquiring 19-month-olds were tested (14 male, 10 female) in Tucson, Arizona. Participants were all of normal birth weight (5lb, 7oz) and term (37 weeks). Data from an additional 17 infants were not included due to experimenter or parental error ($n = 2$) and fussiness ($n = 15$). Note that this fuss-out rate is comparable to that of Experiment 1.

3.1.2 Materials—Materials were designed to be as similar as possible to the English-based materials used in H&G and Experiment 1 of the present paper, but the words and sentences were phonemically, phonotactically, and prosodically based on Japanese. The stimuli were recorded in Tokyo by a young adult native Japanese speaker from Aichi Prefecture with an accent similar to the Tokyo dialect. She was instructed to speak in a happy, infant-directed register. The recorded sentences were partially standardized for amplitude, tempo, and pause length using the criteria and methods discussed in Section 2.1.2.

As in Experiment 1, the familiarization sentences each had six words with either 1-clause (ABCDEF) or 2-clause (ABC, DEF) prosody. The 1-clause sentences were modeled on (1) and the 2-clause sentences were modeled on (2). As with the English stimuli, each word

class had two tokens, for a total of 64 unique word combinations (see Table 3, below), and the word combinations used at test were withheld from familiarization in order to test whether the infants were generalizing at the level of the word class. The non-withheld sentences (25 unique word combinations) were organized into pseudorandomized familiarization lists (1-clause and 2-clause), with each sentence separated by a 1-second pause.

- (1) M ii ame huru e-o mita already good rain fell picture-ACC looked “I’ve already seen the good picture of falling rain.” -- OR -- “I’ve already seen the picture of good falling rain.”
- (2) Ame huru naka, ii e-o mita rain fell while good picture-ACC looked “While it rained, I looked at a good picture.”

While the English-based words were CVC, the Japanese-based words comprised a consonant followed by a long vowel, a diphthong, or a nasal coda consonant. Thus, the English-based and Japanese-based word sets were both bi-moraic in a way that complied with the syllable structure restrictions of the target language. For the English-acquiring participants, the Japanese-based words included several non-native features, including the phonemes [, ,] and the presence of long vowels.

Sentences were analyzed for the presence of clause-level prosodic cues. For the 2-clause items, there was a mean inter-clause pause of 526ms between words C and D. Note that this is longer than the 173ms mean inter-clause pause duration for the English-based stimuli. This is likely due to an overall slower speaking rate for the Japanese speaker (e.g., the clauses ABC and DEF had a mean duration of 994 for the Japanese speaker and 382 for the English speaker). As with the English-based stimuli, there were no measurable pauses in the 1-clause sentences. Word rimes were significantly longer for words at the internal and final clause boundaries than for non-clause-final words ($F(1, 57) = 144.1, p < .001$) (Table 4). The 1-clause and 2-clause sentences both had an appropriate intonational contour for the intended prosody (see Figure 3).

Test items had the same 2-clause prosody as the 2-clause familiarization items. As in Experiment 1, grammatical-movement items were of the form DEF, ABC and ungrammatical-movement items were of the form EFA, BCD. Test sentences were individually recorded to ensure naturalness, and only those items that had unfamiliarized word combinations were included in the final test materials.

3.1.3 Design, and Task—The design and task were identical to H&G and Experiment 1 of the present paper. Infants were familiarized with either 1-clause (ABCDEF) or 2-clause (ABC, DEF) sentences over a an initial long familiarization phase during which they played with toys in a small room. This was followed by an additional short familiarization phase and twelve test trials using the Head-Turn Preference Procedure. The Japanese sentences had a slightly longer duration than their English counterparts, so the familiarization audio files were longer in Experiment 2 to keep the total number of sentences consistent. The initial familiarization lasted 12.62 minutes, and the final familiarization in the testing booth lasted 2.46 minutes.

Grammatical-movement test items (DEF, ABC) involved a reordering of the prosodically-grouped constituents from the 2-clause familiarization items, and ungrammatical-movement items (EFA, BCD) violated the prosodically-marked constituent structure of the 2-clause familiarization items.

3.2 Results

The mean looking time across conditions was 7.02s, once looking times shorter than two seconds were excluded. Nine of the 12 infants from the 2-clause condition and seven of the 12 infants from the 1-clause condition showed a familiarity preference for the grammatical-movement test items.

Looking times were log transformed to increase normality, and the data were analyzed using linear mixed effects modeling, as described in Section 2.2. There was a significant interaction of familiarization (1-clause versus 2-clause) by test condition (grammatical-versus ungrammatical-movement) ($\chi^2(1) = 4.18$, $p = .041$). Infants in the 2-clause condition listened significantly longer to the grammatical-movement test items ($\chi^2(1) = 5.22$, $p = .022$), while the 1-clause participants did not discriminate between the test item types ($\chi^2(1) = .06$, $p = .81$). Unlike in Experiment 1, there were no first trial effects for either the 1-clause or 2-clause group (all $p > .05$)⁴.

3.3 Discussion

Results again support the acoustic salience version of prosodic bootstrapping, bolstering the findings of Experiment 1 and further suggesting that infants do not require extensive prior experience with the prosodic features of the target language to use prosody as a cue to constituent-like units.

Despite the unfamiliar nature of the prosodic boundary cues, English-acquiring infants listening to Japanese-based stimuli produced looking time patterns that were nearly identical to English-acquiring infants listening to English-based stimuli in H&G. This is clearly reflected in Figure 4. In fact, there is no indication that degree of prior experience with the specific instantiations of clause-level prosody in English versus Japanese played any role in the learning that occurred for the American English-acquiring infants.

The similarity between English-acquiring infants' behavior with English and Japanese stimuli is particularly interesting given the extremely different intonational patterns. Seidl (2007), in a replication of Nazzi et al's (2000) experiment on the influence of prosody on memory, found that pitch was the only prosodic cue that was necessary to obtain memory effects in American 6-month-olds – pausing and final-syllable lengthening combined were not sufficient. The present findings suggest that infants are flexible in their expectations of constituent-like intonation. While this type of flexibility would clearly be beneficial to a young infant, who could be born into any language community, the fact that it continues beyond the first year and a half of life suggests that infants remain open-minded about what

⁴Again, the pattern of results is the same when the data are analyzed with ANOVA. There was a significant familiarization by test condition interaction ($F(1,22) = 5.78$, $p = .025$), with a significant effect of test condition for the 2-clause group ($F(1,11) = 7.65$, $p = .018$), but not the 1-clause group ($F < 1$).

constitutes the intonational contour of a clause at least into the beginnings of the syntactically-important two-word stage.

Finally, results bolster a culturally-mediated attentional difference as an explanation for the discrepancy in direction of preference seen in American versus Japanese infants listening to English-based stimuli. Like their compatriots in H&G, participants in the current experiment showed a familiarity preference for the grammatical-movement items. This suggests that some sort of cultural or environmental factor is at play. Note that the infants in the present experiment did, like the American infants in H&G, engage with toys during the initial long familiarization phase. This suggests that the Japanese infants, who often did not play with toys, may indeed have been attending more to the familiarization stimuli playing in the background.

4. General Discussion

Both Japanese and American infants from the 2-clause familiarization groups showed significant discrimination – albeit with different directions of preference – between sentences with movement of prosodically-grouped words versus sentences with movement of word strings that crossed a prosodic boundary, even though the infants were listening to unfamiliar prosody. Coupled with the lack of discrimination in the 1-clause groups, these experiments support the acoustic salience version of the prosodic bootstrapping hypothesis.

To determine how much grammar the infants are learning from non-native prosody, it is useful to consider the current results alongside results from H&G, Experiment 2. It is possible that the discrimination seen in the 2-clause groups of the present paper could be an effect of words in perceptually salient pre- and post-pausal positions, rather than an effect of clausal prosody (which includes pauses, but also pitch and lengthening cues). It could also be the case that the presence of a pause in the 2-clause items of the present experiments facilitates learning of the transitional probabilities between word classes, since it is only necessary to learn the contingency between four syllable pairs for the 2-clause group (A-B, B-C, D-E, and E-F), rather than the five transitions that must be learned by the 1-clause group (A-B, B-C, C-D, D-E, and E-F). However, in H&G, Experiment 2, infants were able to parse 2-clause utterances into constituent-like chunks, even when the clauses were marked with final lengthening and pitch cues, but no pause (for further details and discussion, see Section 2.3). Thus, our present results, considered alongside results from H&G, Experiment 2, suggest that at 19 months, infants are willing to entertain non-native prosody as a useful cue in a grammar learning task. In fact, infants do not seem to perform more poorly with non-native prosody than they did with native prosody in H&G.

There is, however, an alternative way to interpret these results that we have glossed over so far: rather than evidence in favor of the acoustic salience hypothesis, our results may perhaps more accurately be considered evidence against the language-specific hypothesis. Despite many prosodic differences in English and Japanese IPs/clauses, the similarities in prosodic boundary cues across languages – even across such a prosodically-distinct pair – may be sufficient to swamp out the infant's previously acquired knowledge of the prosody-syntax interface in her native language. If this is the case, our results still provide clear

evidence against the language-specific hypothesis, because the infants were able to learn from unfamiliar prosody, despite their 19 months of experience with the prosody of their native language. However, rather than relying purely on the acoustic salience of prosody, the infants could have generalized learned knowledge about prosodic boundary cues in their native language and applied it to the unfamiliar language heard during the experimental session.

Similarly, 19-month-olds have likely already mastered the task of locating major syntactic constituents in their native language, relying on a host of cues, including prosody, grammatical morphemes, word meanings, cross-sentential comparisons, etc. The fact that they can use prosody to locate cohesive, constituent-like chunks in a novel artificial language – with unfamiliar words, prosody, phonemes, and syntactic structure – is consistent with the prerequisite of bootstrapping that the cue be readily available without extensive prior learning. It also suggests that prosody may be a useful bootstrapping tool for internationally-adopted children, who must learn a new language from scratch (c.f. Snedeker, Geren, & Shafto, 2007). Nonetheless, parsing speech into large syntactic constituents like sentences and clauses likely precedes parsing speech into smaller-level units like words, so it must happen early in typical first language acquisition. Our results show that 19-month-olds can use prosody to locate constituent-like groupings – sequences of abstract word classes (A, B, etc.) that can appear as a unit in a new position of the sentence and can carry either sentence-initial or sentence-final prosody – in a novel language. Whether younger infants use prosody to bootstrap knowledge of such constituent-like chunks in their native language remains an open question.

Lending support to the idea that younger infants may use prosody differently than the older infants tested here, recall that Johnson and Seidl (2008) report that English- and Dutch-acquiring 6-month-olds were not able to use prosody from the counterpart language to recognize a previously heard clause that was embedded in a new passage of speech. This is despite the fact that English and Dutch are much more prosodically similar than English and Japanese. It is surprising, then, that the English and Japanese learners in the present experiments could use the prosody of the counterpart language to extract clause-like chunks from an artificial grammar. Differences in materials or task across the two studies could account for this discrepancy. For example, though the infants tested by Johnson and Seidl were too young to understand the sentences, the stimuli did involve real sentences with real words and may, therefore, have been more natural-sounding. However, the strength of the boundary cues in Johnson and Seidl's stimuli appear to be as strong as those in our stimuli, at least for pauses. The intra-clause pause durations in Johnson and Seidl's Dutch stimuli were 630 and 820ms (details are not given about the English stimuli), which are longer than the 173ms (English) and 526ms (Japanese) mean pauses in our stimulus items. It is also possible that there were other aspects of the phonology (phonemes used, phonotactic patterns, etc.) that could have distracted the infants from the prosodic contour.

Alternatively, this pattern of results may suggest a change in how prosody is used across ages and for different tasks. For very young infants, prosody is interpreted merely as an acoustic cue (c.f., Jusczyk, 1989). Somewhat older infants, such as the 6-month-olds tested by Johnson and Seidl, are focused on the prosodic features of their native language and fail

to use non-native prosody to bootstrap syntactic knowledge. By the time infants have reached 19 months, the acoustic salience of non-native prosody is again sufficient to induce prosody-based learning in a new language or, alternatively, infants are able to generalize their knowledge of prosodic boundary cues from their native language to a new language. However, null results from a single experimental paradigm, coupled with the positive results reported here, only begin to touch on the developmental trajectory of prosodic bootstrapping and prosody-based syntactic segmentation with non-native prosody.

Two issues remain that merit comment. First, while we have focused on the acoustic salience and language-specific versions of prosodic bootstrapping as they relate to parsing speech into syntactically-relevant chunks, prosodic bootstrapping also plays a role in word segmentation. In fact, the prosodic cues to word boundaries vary more from language to language than they do for syntactic boundaries. For example, most content words in English begin with a stressed syllable, while Polish words have penultimate stress; in French, stress marks the ends of phrases rather than words, and Greenlandic does not have stress at all. Therefore, it is quite possible that the language-specific version of prosodic bootstrapping holds for word segmentation, while the acoustic salience version obtains for syntactic segmentation. Since parsing clauses likely precedes parsing words, this would make sense. To date, though, this remains an open question. It is also probable that both mechanisms play a role in multiple learning domains, and that experimental evidence so far simply does not paint a complete picture.

Finally, our results touch on the debate about the degree to which syntax acquisition relies on domain-general versus Language-specific learning mechanisms (harkening back to Fodor, 1983). We distinguish here between *language-specific* and *Language-specific* – the former (used throughout the paper) indicating ‘specific to a particular language (and, therefore, learned)’ and the latter indicating ‘specific to the human faculty of language.’ Our results support the acoustic salience version of the prosodic bootstrapping hypothesis, hinting that prosody's role in syntax acquisition could well be due to perceptual grouping mechanisms and reliance on the salient nature of prosodic boundary cues. The infant's language faculty may not come into play at all. However, we have no reason not to believe that our infant participants treated the stimulus sentences as Language, despite the non-native words, prosody, and syntax. Thus, our results suggest that older infants can use prosody to bootstrap into the syntax of an unfamiliar language after only brief exposure to that language. Whether this ability is due to general learning mechanisms or is specific to Language is a matter for future research.

Acknowledgments

This research was supported by National Science Foundation Grant 0950601 and NICHD #R01 HD042170 to LouAnn Gerken and a National Science Foundation Graduate Research Fellowship and an East Asia and Pacific Summer Institute Grant to Kara Hawthorne. Additional funding was provided via RIKEN Brain Science Institute funding to Reiko Mazuka. Thank you to the reviewers for comments on this work and to the families for their participation. I am also grateful to Sara Knight, as well as Omar Aujani, Jeannette Bell, Lauren Clough, Erin Huff, Kelly McGinn, Brianna McMillan, and Juliet Minton from the Tweety Lab at the University of Arizona and Mihoko Hasegawa, Yuri Hatano, Mari Kanamura, Ai Kanato, Yumiko Maeda, and Yuka Miyake at the Laboratory for Language Development at RIKEN Brain Science Institute. Portions of this paper were included in the first author's dissertation.

References

- Baayen, H. languageR: Data sets and functions with “Analyzing linguistic data: A practical introduction to statistics.” R package version 1.4.1. 2014. URL: <http://CRAN.R-project.org/package=languageR>.
- Bates, D.; Maechler, M.; Bolker, B.; Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. 2014. URL: <http://CRAN.R-project.org/package=lme4>.
- Beach CM, Katz WF, Skowronski A. Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America*. 1996; 99(2):1148–1160. [PubMed: 8609299]
- Beckman, ME.; Pierrehumbert, JB. Proceedings of the 24th ACL. New York: 1986b. Japanese prosodic phrasing and intonation synthesis.; p. 173-180.
- Broen PA. The verbal environment of the language-learning child. Monograph of American Speech and Hearing Association. 1972; 17:1–103.
- Caudill W, Weinstein H. Maternal care and infant behavior in Japan and America. *Psychiatry: Journal for the Study of Interpersonal Processes*. 1969; 32:12–43.
- Christophe A, Dupoux E, Bertoncini J, Mehler J. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*. 1994; 95:1570. [PubMed: 8176060]
- Christophe A, Gout A, Peperkamp S, Morgan J. Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*. 2003; 31:585–598.
- Christophe A, Mehler J, Sebastián Gallés N. Perception of prosodic boundary correlates by newborn infants. *Infancy*. 2001; 2:385–394.
- Cooper, WE.; Paccia-Cooper, J. Syntax and speech. Harvard University Press; Cambridge, MA: 1980.
- Cruttenden, A. Intonation. Cambridge University Press; Cambridge, England: 1986.
- Fernald, A.; McRoberts, G. Prosodic bootstrapping: A critical analysis of the argument and the evidence.. In: Morgan, JL.; Demuth, K., editors. Signal to syntax. Erlbaum; Mahway, NJ: 1996. p. 365-388.
- Fernald A, Taeschner T, Dunn J, Papousek M, de Boysson-Bardies B, Fukui I. A cross-linguistic study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*. 1989; 16:477–501. [PubMed: 2808569]
- Fisher C, Tokura H. Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*. 1996; 67:3192–3218. [PubMed: 9071777]
- Fodor, JA. The modularity of mind: An essay on faculty psychology. MIT Press; Cambridge, MA: 1993.
- Friederici AD, Friedrich M, Christophe A. Brain responses in 4-month-old infants are already language specific. *Current Biology*. 2007; 17:1208–1211. [PubMed: 17583508]
- Gleitman, LR.; Wanner, E. Language acquisition: The state of the art.. In: Wanner, E.; Gleitman, LR., editors. Language acquisition: The state of the art. Cambridge University Press; Cambridge: 1982. p. 3-48.
- Gout A, Christophe A, Morgan JL. Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*. 2004; 51:548–567.
- Halle PA, de Boysson-Bardies B, Vihman M. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*. 1991; 34:299–318. [PubMed: 1843528]
- Hawthorne K, Gerken LA. From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*. 2014; 133(2):420–428. [PubMed: 25151251]
- Hirsh-Pasek K, Kemler Nelson DG, Jusczyk PW, Wright Cassidy K, Druss B, Kennedy L. Clauses are perceptual units for young infants. *Cognition*. 1987; 26:268–286.
- Houston-Price C, Nakai S. Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*. 2004; 13:341–348.
- Hunter MA, Ames EW. A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*. 1988; 5:69–95.

- Igarashi Y, Nishikawa K, Tanaka K, Mazuka R. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America*. 2013; 134(2):1283–1294. [PubMed: 23927126]
- Iivonen, A. Intonation in Finnish. Cambridge University Press; Cambridge: 1998. p. 311–327.
- Johnson EK, Seidl A. Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*. 2008; 13:445.
- Jusczyk, PW. Perception of cues to clausal units in native and non-native languages.. Paper presented at the Biennial Meeting of the Society for Research in Child Development; Kansas City, Missouri. 1989.
- Jusczyk PW, Cutler A, Redanz NJ. Infants' preference for the predominant stress patterns of English words. *Child Development*. 1993; 64:675–687. [PubMed: 8339688]
- Jusczyk PW, Kemler Nelson DG, Hirsh-Pasek K, Kennedy L, Woodward A, Piwoz J. Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*. 1992; 24:252–293. [PubMed: 1582173]
- Kawahara S, Shinya T. The intonation of gapping and coordination in Japanese: Evidence for intonational phrase and utterance. *Phonetica*. 2008; 65:62–105. [PubMed: 18523367]
- Kemler Nelson DG, Hirsh-Pasek K, Jusczyk PW, Wright-Cassidy K. How the prosodic cues in motherese might assist language learning. *Journal of Child Language*. 1989; 16:55–68. [PubMed: 2925815]
- Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A, Gerken LA. The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*. 1995; 18:111–116.
- Langus A, Marchetto E, Bion RAH, Nespor M. Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*. 2012; 66:285–306.
- Mandel DR, Jusczyk PW, Kemler Nelson DG. Does sentential prosody help infants organize and remember speech information? *Cognition*. 1994; 53(2):155–180. [PubMed: 7805352]
- Mandel DR, Kemler Nelson DG, Jusczyk PW. Infants remember the order of words in a spoken sentence. *Cognitive Development*. 1996; 11(2):181–196.
- Martin, A.; Igarashi, Y.; Jincho, N.; Reiko, M. Utterances in infant-directed speech are shorter, not slower. under review
- Morgan, JL. From simple input to complex grammar. MIT Press; Cambridge, MA: 1986.
- Nazzi T, Bertoncini J, Mehler J. Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*. 1998; 24:756–766. [PubMed: 9627414]
- Nazzi T, Kemler Nelson DG, Jusczyk PW, Jusczyk AM. Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*. 2000; 1:124–147.
- Nazzi T, Ramus F. Perception and acquisition of linguistic rhythm by infants. *Speech Communication*. 2003; 41(1):233–243.
- Nespor M, Shukla M, Van de Vijver R, Avesani C, Schraudolf H, Donati C. Different phrasal prominence realizations in VO and OV languages. *Lingue E Linguaggio*. 2008; VII.2:1–29.
- Nespor, M.; Vogel, I. Prosodic phonology. 2nd ed. Foris Press; Dordrecht, Holland: 2007.
- Nishinuma, Y. Un modèle d'analyse automatique de la prosodie : Accent et intonation en japonais. Éditions du Centre National de la Recherche Scientifique; Paris: 1979.
- Peters, AM. The units of language acquisition. Cambridge University Press; Cambridge: 1983.
- Pierrehumbert, J. Doctoral Dissertation. Massachusetts Institute of Technology; Cambridge, MA: 1980. The phonology and phonetics of English intonation..
- Pilon R. Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research*. 1981; 10:113–122.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. URL <http://www.R-project.org/>
- Ramus F, Hauser MD, Miller C, Morris D, Mehler J. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*. 2000; 288:349–351. [PubMed: 10764650]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274(5294):1926–1928. [PubMed: 8943209]

- Seidl A. Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*. 2007; 57:24–48.
- Seidl A, Hollich G, Jusczyk PW. Early understanding of subject and object wh-questions. *Infancy*. 2003; 4(3):423–435.
- Shukla M, White KS, Aslin RN. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*. 2011; 108(15): 6038–6043.
- Snedeker J, Geren J, Shafto CL. Starting over: International adoption as a natural experiment in language development. *Psychological Science*. 2007; 18(1):79–87. [PubMed: 17362382]
- Snedeker J, Yuan S. Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*. 2008; 58(2):574–608. [PubMed: 19190721]
- Soderstrom M, Kemler Nelson DG, Jusczyk PW. Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior & Development*. 2005; 28:87–94.
- Soderstrom M, Morgan JL. Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science*. 2007; 10(5):641–653. [PubMed: 17683348]
- Soderstrom M, Seidl A, Kemler Nelson DG, Juscayk PW. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*. 2003; 49:249–267.
- Soderstrom M, Blossom M, Foygel R, Morgan JL. Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*. 2008; 35:869–902. [PubMed: 18838016]
- Takeda K, Sagisaka Y, Kuwabara H. On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America*. 1989; 86:2081–2096.
- Toro JM, Trobalon JB, Sebastián-Gallés N. The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*. 2003; 6:131–136. [PubMed: 12728358]
- Ueyama M. An experimental study of vowel duration in phrase-final contexts in Japanese. *UCLA Working Papers in Phonetics*. 1999:174–182.
- Venditti, JJ.; Jun, SA.; Beckman, ME. Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English.. In: Morgan, JL.; Demuth, K., editors. *Signal to syntax*. Erlbaum; Mahway, NJ: 1996. p. 287-311.
- Venditti, JJ.; Yamashita, H. Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP). Vol. 3. Yokohama, Japan: 1994. Prosodic information and processing of temporarily ambiguous constructions in Japanese.; p. 1147-1150.
- Wakefield JA, Doughtie EB, Yom BH. The identification of structural components of an unknown language. *Journal of Psycholinguistic Research*. 1974; 3:261–269.
- Yoshida KA, Iversen JR, Patel AD, Mazuka R, Nito H, Gervain J, Werker JF. The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*. 2010; 115:256–361.

Highlights

Japanese infants listened to an English-based artificial grammar and vice-versa.

Both groups treated prosodically-grouped words as more constituent-like at test.

Infants can use either native or non-native prosody for this task.

Prior learning of specific correlations between syntax and prosody is not necessary.

Prosody is useful in syntax acquisition because it is acoustically salient.

FIGURE 1.

Example waveforms and pitch contours for a 2-clause and 1-clause English-based familiarization sentence.

FIGURE 2.

Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions.

FIGURE 3.
Example waveforms and pitch contours for a 2-clause and 1-clause Japanese-based familiarization sentence.

FIGURE 4.

Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions.

TABLE 1

English stimuli words (in IPA).

		Word Classes					
		<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>
token 1		b	d	kæg	f	b	h s
		p	v				k
token 2		n	p	l	t	z	zæf v t
		m		d			

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Duration measurements represent the total vowel duration, exclusive of the onset and coda duration.

	Mean Duration (ms)						
	<u>A</u>	<u>B</u>	<u>C</u>	<u>pause</u>	<u>D</u>	<u>E</u>	<u>F</u>
2-clause familiarization items	80	104	207	173	122	104	146
1-clause familiarization items	76	88	82	n/a	122	121	135

TABLE 3

Japanese stimuli words (in IPA)

Word Classes						
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>
token 1	gi:	a:		ai	po:	ne:
token 2	eŋ	oi	s	ŋ	toŋ	jŋ
	doi					

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 4

Duration measurements for words represent the total rime duration.

	Mean Duration (ms)						
	<u>A</u>	<u>B</u>	<u>C</u>	<u>pause</u>	<u>D</u>	<u>E</u>	<u>F</u>
2-clause familiarization items	310	262	420	526	247	271	477
1-clause familiarization items	283	239	291	n/a	243	271	415

Note that the English duration values in Table 2 represent only vowel duration, so the values are not directly comparable. Rime duration was used for the Japanese stimuli because of the presence of long vowels and diphthongs.