Journal of Child Language (2024), **51**, 137–167
doi:10.1017/S0305000922000460

ARTICLE

Utterance-Initial Prosodic Differences Between Statements and Questions in Infant-Directed Speech

Susan GEFFEN[1,3], Kelly BURKINSHAW[2,3], Angeliki ATHANASOPOULOU[2] and Suzanne CURTIN[4,*]

[1]Department of Psychology, Occidental College
[2]School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Canada
[3]Department of Psychology, University of Calgary, Canada
[4]Department of Child and Youth Studies, Brock University, Canada
*Corresponding author: Suzanne Curtin, Ph.D. Department of Child and Youth Studies Brock University, Canada. E-mail: scurtin@brocku.ca

(Received 16 April 2021; revised 15 July 2022; accepted 19 August 2022)

Abstract

Cross-linguistically, statements and questions broadly differ in syntactic organization. To learn the syntactic properties of each sentence type, learners might first rely on non-syntactic information. This paper analyzed prosodic differences between infant-directed *wh*-questions and statements to determine what kinds of cues might be available. We predicted there would be a significant difference depending on the first words that appear in *wh*-questions (e.g., two closed-class words; meaning words from a category that rarely changes) compared to the variety of first words found in statements. We measured F0, duration, and intensity of the first two words in statements and *wh*-questions in naturalistic speech from 13 mother-child dyads in the Brent corpus of the CHILDES database. Results found larger differences between sentence-types when the second word was an open-class not a closed-class word, suggesting a relationship between prosodic and syntactic information in an utterance-initial position that infants may use to make sentence-type distinctions.

Keywords: prosody; statements; *wh*- questions; input; and infant-directed speech

Introduction

Distinguishing between statements and questions is an important part of syntactic acquisition. In English, statements and questions differ both prosodically and syntactically. For example, the canonical ordering in simple transitive declarative sentences is subject-verb-object (Slobin & Bever, 1982), as in *Anna likes chocolate* (examples brought in and expanded from Geffen & Mintz, 2017). In comparison, (yes/no) questions are often characterized by auxiliary-inversion and do-support (e.g., *Anna does like chocolate* becomes *Does Anna like chocolate?*). *Wh*- questions additionally include an initial *wh*-word such as *who, what, how* (e.g., *What does Anna like?*). Previous studies have found that questions account for approximately half of the utterances in young children's input with yes/no questions comprising 23% of total utterances, and *wh-* questions 21% (Newport, 1977). Given the differences in word order between statements and questions, the ability to distinguish between statements and questions could give learners a basis from which to develop an analysis of differing grammatical structures (Slobin & Bever, 1982). However, it is unlikely that infants will make initial sentence-type distinctions based on word order.

While infants need to acquire knowledge of the grammatical structures of statements and questions more generally, *wh-* questions provide an interesting challenge for young language learners. They are syntactically similar to yes/no questions (e.g., AUX inversion), yet they are prosodically similar to statements, typically ending with final flat or falling intonation (Bartels, 1999; Hedberg, Sosa & Fadden, 2004). This raises the question of how infants learn to distinguish between statements and *wh-* questions. Another question is whether infants use the same strategy to distinguish statements from all types of questions, or whether infants must rely on different strategies for distinguishing between different types of questions (e.g., yes/no versus *wh-* questions). While both questions are of interest to the field, this paper will focus on the first question.

Prosody as a Cue to Sentence Type: Pitch

One possible cue to the distinction between statements and questions is prosody. Prosody is realized as changes in pitch, duration, and intensity. Pitch is arguably the main prosodic cue for distinguishing sentence types. English relies on the final pitch contour as an important perceptual cue for sentence-type discrimination (e.g., American English – Săfářǒvá & Swerts, 2004; British English [e.g., London, Cambridge] – Grabe, 2004), as do many other languages (e.g., Castilian Spanish – Face, 2007; French – Vion & Colas, 2006). In adult-directed speech, yes/no questions typically end with a final rise in pitch, while statements end with final flat or falling intonation (Bartels, 1999; Hedberg et al., 2004; Ladd, 2008) with a few exceptions (e.g., in Belfast English, both statements and questions have rising intonation – – Grabe, 2004; Jarman & Cruttenden, 1976). For example, American English speakers were more likely to identify utterances as declarative questions if the utterances ended with a final rise (Săfářǒvá & Swerts, 2004). The same is true of declarative questions, which differ from declarative statements only in prosody (e.g., *Anna likes chocolate?*). *Wh-* questions usually have a falling or level pitch, and thus have final contours similar to statements (Hedberg et al., 2004; Ladd, 2008; Ladefoged & Johnson, 2010). The contours described above (apart from the distinction between declarative statements and questions) have also been found in infant-directed speech (IDS) (Geffen & Mintz, 2017).

As stated above, the final pitch contour is a critical region for discriminating between questions and statements. But what about when the final pitch contour is not distinct across utterance types, as with statements versus *wh*-questions (e.g., Geffen & Mintz, 2017)? If prosody is an important cue for understanding this distinction, then it could come into play elsewhere in the sentence. Although most studies that have examined the acoustic features of questions have focused on the ends of utterances, work by O'Shaughnessy (1979) examined the first, medial and last accented syllable of yes/no questions and statements. The findings demonstrated that the question intonation affects the entire fundamental frequency (F0) contour and is not limited to the final rise or fall at the end of an utterance. The O'Shaughnessy (1979) study suggests that prosodic information is available in multiple places across the utterance although the end of the utterance is the most common location (e.g., Bartels, 1999; Hedberg et al., 2004; Ladd, 2008; Ladefoged & Johnson, 2010). More recently, a production experiment shows utterance-initial differences between Canadian English statements and both yes/no and declarative questions (although no differences between the two question types), with statements demonstrating a higher initial pitch accent, earlier pitch peak alignment and smaller F0 change (Patience, Marasco, Colanton, Klassen, Radu & Tararova, 2018). These studies show that prosodic cues have been observed in multiple locations throughout the sentence in the production of sentence types, and in some instances can be used to distinguish between statements and questions. Thus, it is important to evaluate as many sentence types as possible, including different types of statements and questions to determine whether prosodic information can help to disambiguate between them.

Although most of the studies we have discussed so far have focused on statements versus yes/no questions, there have been several studies that have specifically looked at *wh-* questions. Hedberg and Sosa (2002) found that *wh-* words were often marked with a rising pitch peak accent in *wh-* questions, as was the fronted negative auxiliary in negative yes/no questions (e.g., *Isn't that kind of underhanded?)*, suggesting there may be a common interrogative marker at the beginning of utterances, although Hedberg, Sosa, Görgülü, and Mameni (2010) cautioned that this effect may have arisen from the speech pattern of an individual speaker. Maxwell and Fletcher (2013) found similar results in Bengali English, in which adult speakers frequently produced a rising pitch peak accent on *wh-* words in *wh-* questions. Hedberg et al. (2010) found that American English *wh-* questions frequently had nuclear accents on either the *wh-* word or the immediately following auxiliary.

Perceptual studies have also demonstrated that adults are sensitive to these prosodic cues for making sentence-type distinctions. Most languages rely on some form of pitch cue to distinguish between statements and questions (e.g., Gussenhoven & Chen, 2000). For example, in languages like English and French that use lexical markings (e.g., auxiliary verbs) to make sentence-type distinctions, adults primarily rely on fundamental frequency for perceiving distinctions between sentence types (English – Cruttenden, 1986; Lieberman, 1967; French – Vion & Colas, 2006), although how F0 changes depends on the language. For example, gating studies with adults have demonstrated greatest accuracy in sentence type discrimination when the final rise is present or absent (Dutch – van Heuven & Haan, 2000; French – Vion & Colas, 2006), though accuracy begins to increase more rapidly roughly halfway through the sentence (e.g., 60-65% of Dutch listeners could accurately identify sentence types when a phrase was truncated right before the second accent and increased to 80% accuracy when the phrase included the second accent; van Heuven & Haan, 2000). Recent perception studies demonstrate that the prenuclear region (often the first word) provides sufficient cues (e.g., pitch accent height, pitch peak alignment timing and F0 change; Patience et al., 2018) for adult listeners to distinguish between statements and questions (Canadian English – Saindon, Trehub, Schellenberg & van Lieshout, 2017; German – Petrone & Niebuhr, 2014). For example, a gating study found that 18% of Canadian English adult listeners identified questions at above-chance levels after hearing only a single word and increased to 49% of listeners after hearing three words (of a five-word utterance; Saindon et al., 2017). A study with German adult listeners found that participants were better at matching utterances with question contexts when prenuclear accents had shallower slopes (20-28.2% for steeper slopes, 81.2-83.5% for shallower slopes; Petrone & Niebuhr, 2014). These results may be partly attributable to cue weighting and lexical knowledge (which infants do not have). However, the specific cues and timing vary between languages. In Dutch, the *wh-* word in *wh*-questions is often characterized by a pitch accent (van Heuven & Haan, 2000). In European Portuguese, participants can correctly identify statements as early as the first stressed vowel but cannot correctly identify declarative questions until the penultimate or final stressed vowel (although this was not the case in sentences that began with *wh-* words; Falé & Faria, 2006).

The question remains whether other prosodic cues could be used to distinguish between statements and questions, as well as whether prosodic patterns earlier in the utterance can be used to distinguish between yes/no and *wh*-questions. The Geffen and Mintz (2017) study suggests that prosodic cues at the end of the utterance are not sufficient for making this distinction in infant-directed speech. However, a recent study by Chiang, Geffen, and Mintz (2018) using the same IDS corpus found that prosodic cues (primarily pitch) in the first two syllables of utterances did differ between statements and questions but did not differ between yes/no and *wh-* questions. This suggests that sentence-initial prosodic cues are available for distinguishing between statements and *wh-* questions in infant-directed speech, although it remains to be seen whether infants are sensitive to these prosodic cues.

If infant-directed speech is organized similarly to adult-directed speech, prosody could provide a robust tool for infants to initially distinguish sentence types. Unlike adults, who can use lexical knowledge to distinguish between sentence types, infants who are only beginning to recognize words and do not have the lexical or syntactic knowledge necessary to distinguish between statements and questions may be able to use prosodic cues to begin to distinguish between them. Indeed, Geffen and Mintz (2017) suggested that infants may initially distinguish between statements and yes/no questions prosodically, allowing them to recognize the distributional similarity between yes/no and *wh-* questions. Soderstrom, Ko, and Nevzorova (2011) found that English-learning children could distinguish declarative statements from declarative questions (questions that have statement word order with rising pitch, e.g., *Anna likes chocolate?*) by two years of age. However, the wide age range of the subjects (between 4.5 months and 2;0 years) in the Soderstrom et al. (2011) study makes it difficult to determine when this ability emerges.

While it remains unclear when English-learning infants begin to use prosodic information to distinguish between sentence types, there is a wealth of evidence demonstrating infants' general sensitivity to prosodic information from an early age. For example, infants can use prosody to make distinctions between broad rhythmic classes of languages as young as two days old (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison, 1988; Moon, Cooper & Fifer, 1993; Nazzi, Bertoncini & Mehler, 1998; see Nespor, Shukla & Mehler, 2011, for an overview of rhythmic classification). Thus, there is ample evidence that infants are sensitive to prosody, suggesting that they could, in principle, use prosodic information to make initial sentence-type distinctions, even if they do not have a specific understanding of what a "statement" or "question" is.

Previous studies (e.g., Chiang et al., 2018; Geffen & Mintz, 2017) have examined whether there are prosodic differences between sentence types without taking word category into account. For example, there are prosodic differences between open-class (e.g., content) and closed-class (e.g., function) words (e.g., Monaghan, Christiansen & Chater, 2007). Given that statements and questions often begin with words that differ in category and frequency, it is important to consider not just whether prosodic information is available to distinguish between sentence types, but also how and whether it is affected

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

by the syntactic categories (e.g., open- versus closed-class word) that characterize the words at the beginning of the utterance.

Duration

While F0 is a driving factor for the discrimination of sentence types, other prosodic cues also play a role. Duration is a secondary prosodic correlate of sentence type in other Germanic languages (e.g., Dutch, Orkney English – van Heuven & van Zanten, 2005) in which pitch is the primary prosodic cue. Perceptual studies have demonstrated that infants can use prosodic cues including duration to identify syntactic unit boundaries such as clauses (e.g., Hirsh-Pasek, Kemler Nelson, Jusczyk, Cassidy, Druss & Kennedy, 1987; Kemler Nelson, Hirsh-Pasek, Jusczyk & Cassidy, 1989) and phrases (Gerken, Jusczyk & Mandel, 1994; Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward & Piwoz, 1992). 9-month-olds can use syllable duration differences (e.g., "longer duration of the syllable immediately preceding a major phrasal boundary"; Jusczyk et al., 1992, p. 289) to detect syntactic boundaries, but only in conjunction with other converging cues such as pitch changes (e.g., Hirsh-Pasek et al., 1987; Jusczyk et al., 1992). Thus, duration, in conjunction with pitch, could help to discriminate statements and questions, but it remains to be seen whether it is useful as an independent cue for sentence-type discrimination. Geffen and Mintz (2017) found that final syllable duration did differ between infant-directed statements and yes/no questions, but not between statements and *wh*-questions. This difference could be attributed to the last word being one syllable in most statements but two syllables in at least half the yes/no questions, suggesting the difference exists at the word (category) level rather than the sentence level. Patel and Grigos (2006) found similar final vowel duration differences when 4-, 7- and 11-year-olds produced declarative questions and statements, although 4-year-olds relied most heavily on final syllable lengthening, while 7- and 11-year-olds used a combination of final syllable lengthening and F0 to indicate sentence-type contrasts. The Patel and Grigos (2006) study avoided the potential confound of different syllable lengths in the final word found in Geffen and Mintz (2017) by having children produce the same sentence (consisting of four monosyllabic words) with statement and question intonation. It is possible the same pattern of final vowel duration differences will be found in the initial words of statements and *wh*-questions – although it seems unlikely given that many statements begin with determiners or pronouns; and *wh*- questions, by definition, begin with one syllable *wh*-words. Alternatively, research shows there are prosodic differences between open-class and closed-class words (Monaghan et al., 2007), including overall duration, so duration may still be a useful cue given that all *wh*- questions will start with a (closed-class) *wh*-word, but statements in an infant's daily input could begin with a greater variety of words including closed-class (e.g., determiners, pronouns) and open-class (e.g., nouns) words.

Intensity

Intensity correlates with the other dimensions of prosody, especially pitch, and infants appear to be sensitive to this correlation (Fernald, 1984). For example, in English, stressed syllables are marked by higher pitch and intensity compared to non-stressed syllables (e.g., Fry, 1955; Hay & Saffran, 2012). There are significant differences in intensity between statements and *wh*- questions in infant-directed speech, with statements demonstrating significantly higher intensity in both the penultimate and final syllables

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

(Geffen & Mintz, 2017). Thus, as with duration, to the extent that intensity correlates with other prosodic cues that differentiate statements from questions, intensity could contribute to sentence-type discrimination.

Acquisition of sentence types and word categories

In English, and in many other languages, words can be separated into two broad classes of words: open-class words and closed-class words. These two classes of words differ in their acoustic and distributional properties. Open-class words include nouns, verbs, and adjectives. Closed-class words provide information about the grammatical relationships between words (e.g., articles, prepositions; Morgan, Shi & Allopenna, 1996). Compared to open-class words, closed-class words are typically characterized by syllable reduction, reduced vowels, and simplified syllable structure with minimal, if any, onsets and codas. These word classes can be distinguished by surface acoustic and phonological information cross-linguistically (English, Dutch, and French – Monaghan et al., 2007; Mandarin and Turkish – Shi, Werker & Morgan, 1999). Open-class and closed-class words also differ in their frequency in daily speech (Morgan et al., 1996; Shi, Cutler, Werker & Cruickshank, 2006a; Shi & Lepage, 2008).

Infants' sensitivity to the distinction between open-class and closed-class words begins early in development. Newborns distinguish between open-class and closed-class words based on surface acoustic and phonological cues (Shi, Morgan & Allopenna, 1999) and by 6 months, infants prefer listening to open-class words over closed-class words (Shi & Werker, 2001). By 10.5 months, infants show sensitivity to the phonological properties that are characteristic of closed-class words, demonstrating a preference for actual closed-class words rather than nonsense stressed syllables (Shady, 1996). Twelve and 17-month-olds will map a novel pseudo open-class but not a novel pseudo closed-class word to an object (Hochmann, Endress & Mehler, 2010; MacKenzie, Curtin & Graham, 2012). Sixteen-month-olds are sensitive to the position of closed-class words in a sentence and their relations to nouns and verbs (Shady, 1996), and 17-month-olds can use this distributional information to identify closed-class words (Hochmann et al., 2010).

By the first year, infants learning a variety of languages can segment closed-class words from continuous speech (e.g., English – Shi et al., 2006a; German – Höhle & Weissenborn, 2003; French – Shi, Marquis & Gauthier, 2006b). Although closed-class words have higher frequency in speech than open-class words, there is variability in the frequency of different closed-class words. For example, Shi and colleagues evaluated a sample of 290,094 words from the Brent Corpus (Brent & Siskind, 2001) and found that *the* appeared 8513 times while *her* appeared 307 times (Shi et al., 2006a). They then tested whether 8- and 11-month-old infants would show longer looking time to pseudo-nouns that had been preceded by the real closed-class words *the* or *her*, compared to pseudo-nouns that had been preceded by the pseudo-closed-class words *kuh or ler.* In the real closed-class word condition, infants segmented those pseudo-nouns preceded by *the.* They also found a developmental progression, likely shaped by language experience, where 8- but not 11-month-olds also looked longer to pseudo-nouns preceded by *kuh,* but not *ler* (Shi et al., 2006a). This work suggests that the high frequency of the closed-class words can help young language learners to segment novel forms, and this ability is shaped by experience with the language.

Evaluating cues at the word level entails examining prosodic and phonetic differences between word categories (i.e., open- and closed-class words). It is also important to

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

consider when and how infants begin to acquire these broad categories (open- and closed-class) as well as more specific categories (e.g., auxiliary verbs and *wh-* words within the closed-class word category). Rowland, Pine, Lieven, and Theakston (2003) provided evidence that order of acquisition of *wh-* words is correlated with input frequency, i.e., high-frequency *wh-* words (e.g., *what*) are acquired before lower frequency words (e.g., *how*). In addition, *wh-* identity questions (*what is that?*) are often shorter and simpler, which may make them easier to acquire. Seidl, Hollich, and Jusczyk (2003) found that by 15 months, infants demonstrate understanding of subject *wh*-questions, but they do not demonstrate understanding of object *wh-* questions until 20 months. This suggests that word category and input frequency can influence understanding and acquisition of categories at the word- and sentence-level. While *wh-* words are a relevant category when evaluating the beginning of *wh-* questions, it is also important to consider the broad categories that frequently make up the initial words in statements (i.e., open- and closed-class words).

These previous studies suggest that infants are sensitive to the difference between open- and closed-class words from an early age. Given the different types of words that characterize the beginning of statements and *wh-* questions, this knowledge could be useful to infants for making initial sentence-type distinctions.

Outline of the Study

While previous studies (e.g., Chiang et al., 2018; Geffen & Mintz, 2017) have provided preliminary evidence that prosodic information is available at the beginning and end of infant-directed sentences to distinguish between statements and questions, these studies examined prosodic information regardless of syntactic information (e.g., word category). Given that *wh-* questions frequently start with a high frequency, closed-class word (though they may not show the typical acoustic patterns characteristic of other closed-class words, e.g., unstressed) while statements can begin with a closed-class or an open-class word, it is important to consider how prosodic differences across word types interact to aid discrimination. In line with previous research (e.g., Morgan et al., 1996), we classified *wh-* words as closed-class words. Therefore, this paper will focus on the distinction between high-frequency, closed-class (e.g., *wh-* words, pronouns, determiners) and low-frequency, open-class words (e.g., nouns, verbs). We will evaluate specific cues at the word level to determine what cue or combinations of cues are driving infants' initial sentence-type discrimination (prosody, phonetic properties, or a combination of the two).

The current study evaluates whether there are utterance-initial prosodic cues (e.g., higher pitch on the first word relative to the second word) that distinguish between *wh-* questions and different types of statements (e.g., those that begin with closed- versus open-class words) in American English infant-directed speech. The current study aims to replicate the findings of Chiang et al. (2018) by examining typical prosodic contours in a larger corpus, including the speech of 13 mothers to their preverbal infants. In addition, we will extend the findings by taking syntactic categories into account, looking at sentences that begin with open- versus closed-class words, and analyzing the prosodic characteristics at both the phrasal and word level.

We examined prosodic measures in *wh-* questions (closed-class – closed-class) and different types of statements (closed-class – closed-class, closed-class – open-class, open-class – closed-class, open-class – open-class). We analyzed the distribution of pitch (F0), duration, and intensity over the first two words of statements and *wh-* questions in infant-directed speech. We hypothesized that there would be utterance-initial prosodic differences between *wh-* questions and different types of statements in American English infant-directed speech depending on the first words (closed-class) that appear in *wh-* questions compared to the variety of first words found in statements (open- and closed-class). We further predicted that these prosodic and word category cues would serve to correctly classify utterances as either statements or *wh-* questions in the models discussed in the Results section, though we are not claiming that infants understand what statements or questions are. Together, these findings have implications for infants' sentence-type discrimination ability, which we will address in the Discussion.

Methods

Input corpora

We evaluated input to 13 American English-learning children from the Brent corpus (Brent & Siskind, 2001, all dyads except d1, m2 and w1[1,2]) from the CHILDES database (MacWhinney, 2000) for the analysis procedure. We selected the Brent corpus because it contains speech to preverbal infants, and, at the time we conducted the study, provided the largest number of audio recordings of Standard American English-speaking mother-child dyads of any corpora in the database. We analyzed utterances of the mother directed towards her preverbal infant, taken from 1-4[3] sessions spanning an approximately 2-4-week period depending on the subject. Infants' ages ranged from 8.27 to 10.10 months. Recordings were made in the subjects' homes. Mothers were fitted with a small waist pack and a lapel microphone located just below their mouths and instructed to go about their daily routines while home alone with the child (Brent & Siskind, 2001).

Selection of utterances

We analyzed the first two words in 3315 sentences that began with two monosyllabic words. This allowed us to replicate Chiang et al.'s (2018) analysis of the first two syllables in IDS statements and *wh-* questions and address the potential confound of different numbers of syllables in initial words between the two sentence types in Geffen and Mintz (2017). All the *wh-* questions (427) began with two closed-class words[4]. We divided statements into four categories: closed-class – closed-class (cc) (923), closed-class – open-class (co) (773), open-class – closed-class (oc) (931) and open-class – open-class (oo) (261). See Table 1 for the number of utterances per sentence type (statement and *wh*-question) and per phrase type within statements (cc, co, oc, oo) for each speaker, as well as the overall total which ranged from 29-664 utterances. Sentences had to meet

[1]d1, m2 and w1 are participant ID codes assigned to mother-child dyads in the Brent corpus.

[2]We excluded sessions with poor audio quality as these would make it challenging to extract acoustic information (d1), those that did not include any *wh-* questions (m2) and sessions that did not include all four types of statements (w1 did not include any statements of the form closed-class—open-class words).

[3]We only included one session each from four speakers (c1, f1, s3 and w3). Speaker i1 provided the most data with four usable sessions.

[4]There were only 16 *wh-* questions that did not have the closed-class—closed-class structure (they all had the closed-class—open-class structure). We chose to exclude these sentences as there were not enough to draw solid conclusions.

Table 1. Frequency of Utterance Types

Speaker	STATEMENTS cc	co	oc	oo	Wh-QUESTIONS cc	TOTAL
c1	14	5	5	3	2	29
f1	7	11	10	4	13	45
f2	112	117	154	45	64	492
i1	221	143	180	64	56	664
j1	41	36	48	6	55	186
q1	27	45	68	19	13	172
s1	120	61	144	42	48	415
s2	80	54	73	17	43	267
s3	8	14	59	7	31	119
t1	47	34	22	11	20	134
v1	127	153	78	27	46	431
v2	93	71	52	5	31	252
w3	26	29	38	11	5	109
TOTAL	923	773	931	261	427	3315

Note. cc = sentences that start with two closed-class words, co = sentences that start with a closed-class word followed by an open-class word, oc = sentences that start with an open-class word followed by a closed-class word, oo = sentences that start with two open-class words.

several criteria for inclusion (see Appendix A for a full list of selection criteria). We based our initial selection criteria on structural properties. *Wh-* questions were characterized by the typical subject and auxiliary inversion structure of yes/no questions with a *wh*-word such as *who, what, where* or *how* in the utterance initial position, possibly in combination with do-support. We excluded utterances that began with discourse markers (e.g., yeah, oh) before the *wh-* word to ensure consistency in the analyses. We also excluded utterances that began with *wh-* words but were statements (e.g., *what a good girl*) because they did not demonstrate the auxiliary inversion characteristic of *wh-* questions. Finally, we excluded questions that began with a closed-class word followed by an open-class word since there were not enough data points (16 sentences) to be added to the analysis. Statements followed the canonical transitive word order and did not have question intonation (i.e., declarative questions). We did not include statements that only consisted of proper names (e.g., naming phrases [*Big Bird*, *Chips Ahoy*]), although we did include an utterance if the proper name was used in a sentence (e.g., *Hi Pooh.*).

More generally, we did not include utterances that only consisted of a single word (e.g., *What? Yeah*) because we wanted to be able to examine prosodic change over at least two (monosyllabic) words. We also did not include any utterances that contained partial and repeated initial words (e.g., stuttering) because this might not be representative of typical prosodic patterns.

We narrowed our selection further by listening to the accompanying waveform. Like Geffen and Mintz (2017), utterances were excluded if they included vocalization or background noise, making analysis difficult or impossible (e.g., laughing, crying, blowing

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

raspberries). We excluded unintelligible sentences because we used the mother's pronunciation to help define word boundaries. We also excluded utterances that reflected read or rehearsed speech (e.g., songs, reading, reciting the alphabet, etc.) since we were primarily interested in spontaneous speech.

Our analysis focused on word-level comparisons, including duplicate two-word phrases. For example, in the utterances "*let's go in here*" and "*let's go see*"[5]*;* the phrase "let's go" is the target in both. We included duplicate phrases for both statements and *wh*-questions for several reasons. First, Geffen and Mintz (2017) found that including duplicate utterances did not change the results, so we decided to include duplicate two-word phrases for a larger corpus to provide more tokens on which the model can base its analysis (as described in the Results section). (Out of the 3,315 two-word phrases we included in the corpus analysis, 2,249 were duplicates [e.g., "let's go"]). Second, previous studies (e.g., Fernald & Morikawa, 1993; Stern, Spieker, Barnett & MacKain, 1983) find that utterance repetition peaks between 4 and 6 months (up to 20% of maternal utterances are exact repetitions, McRoberts, McDonough & Lakusta, 2009), tapering off to adult-like (i.e., almost nonexistent) levels by 2;0 years. This demonstrates that repetition is a common characteristic of infant-directed speech (e.g., Stern et al., 1983), therefore making it an important part of any evaluation of naturalistic infant-directed speech. Third, the sessions we evaluated from the Brent corpus consist of naturalistic interactions between mothers and their 8- to 10-month-old infants, so we cannot control for the amount of exact or partially replicated phrases or utterances. This is particularly relevant for *wh*- questions since there are only a limited number of *wh*- words, guaranteeing repetitions, especially in the first two words. For example, our corpus included 67 repetitions of "where are" and 207 repetitions of "come on". McRoberts et al. (2009) suggest that repeated utterances can serve as an important framework for infants to notice salient perceptual cues in the speech signal. Thus, including replication provides a more accurate representation of daily input to young infants and may prove useful for syntactic development.

Location of analyses within sentences

Separate analyses of each acoustic measure are detailed below. We analyzed sentences to determine whether there were prosodic differences between *wh*- questions and different types of statements, such as acoustic prominence (e.g., higher pitch) at the beginning of the utterance. Given that previous research (Hedberg & Sosa, 2002; Maxwell & Fletcher, 2013) found rising pitch contour accent on *wh*- words, we reasoned that the first two words (which often correspond to the first two syllables in the utterance, especially in *wh*-questions) were likely to contain prosodic cues that would allow infants to distinguish between *wh*- questions and statements with different phrase structures (e.g., those that begin with closed-class versus open-class words).

While Chiang et al. (2018) found that there were prosodic differences between statements and *wh*- questions, they only analyzed a small number of sentences (roughly 100 utterances of each type). Their study also grouped all types of statements together, irrespective of word category. Here we examine a larger corpus (approximately 3300 phrases) to provide more tokens on which the model can base its analysis and evaluate whether there is differential acoustic prosody on the initial words of *wh*- questions and

[5]From dyad i1, session 0914.

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

different types of statements (e.g., those that begin with closed-class versus open-class words). For each utterance, a single coder marked the boundaries for the first two words for analysis by hand, although multiple coders might have worked on different utterances within the same session. Coders initially marked word boundaries by examining the spectrogram and waveform and listening to the corresponding audio. In addition, there can be a lot of variability between speakers in the combination of syntactic and prosodic cues, especially in infant-directed speech. Therefore, we allowed our coders to use their best judgment as native English speakers (following similar transcription techniques to those used by Bergelson, Casillas, Soderstrom, Seidl, Warlaumont & Amatuni, 2019 and Soderstrom, Blossom, Foygel & Morgan, 2008). The second author performed spot checks on the coding to check for consistency across coders and performed additional checks if multiple problems were found within a session. Coders labeled each section with a rough English transcription of the word. We used a modified version of Lennes' (2003) Praat script (Boersma & Weenink, 2011) to carry out batch extraction of mean F0, maximum F0, minimum F0, F0 range, duration, and mean intensity from each labeled interval (i.e., each of the first two words of each utterance) within all 29 sessions (using the default settings, *F0 range*: 100-500 Hz, *F0 sample rate*: 100 Hz.

Pitch variables

To assess the availability of pitch as a cue to sentence type in infant-directed speech, we measured average pitch (Mean F0) as well as the lowest and highest pitch (Min F0 and Max F0, respectively).

For Mean F0 and Max F0, we were interested primarily in the change of these values across the first two words (see Methods). Mean F0 provides a simple, coarse measure that can be used to identify the general degree of sentence-initial pitch rise or fall. Max F0, while clearly related to Mean F0, provides specific information about pitch peak. Languages like English, which do use lexical markings (e.g., auxiliary verbs) to distinguish sentence types, demonstrate differences in pitch peak location between sentence types. For example, "English *wh-* questions show an earlier pitch peak and final F0 decline" (Best, Levitt & McRoberts, 1991, p. 162). Thus, the location and degree of Max F0 could be a useful cue for distinguishing between sentence types – for the purposes of the current study, we measured Max F0 in each of the words and identified which of the two was higher; we used this (Max F0 for the pair) and noted whether the location of Max F0 (for the pair of words) occurred in the first or second word. Since infant-directed speech is characterized by exaggerated prosodic contours, including higher mean pitch and expanded pitch range (Fernald & Kuhl, 1987), it is an open question which properties of pitch contours that distinguish between sentence types in adult-directed speech are maintained in speech to infants. Geffen and Mintz (2017) found that prosodic cues, primarily pitch, at the end of utterances distinguish statements from yes/no questions but not from *wh-* questions. Chiang et al. (2018) found that similar prosodic cues at the beginning of utterances distinguish statements from yes/no and *wh-* questions, but do not distinguish these question types from each other. These results suggest that prosodic information is available at various points in an utterance to distinguish statements from different types of questions, but there does not appear to be one consistent set of prosodic cues by itself (e.g., pitch, intensity), that distinguishes all three sentence types.

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

For analyzing PITCH RANGE at the word level, we subtracted the F0 min from the F0 max regardless of their location in the word. For example, if the F0 min was 150.1 Hz and the F0 max was 390.9 Hz, the pitch range would be 240.8 Hz (390.9 - 150.1).

Duration and intensity

Duration and average intensity were extracted from the marked words in Praat (Boersma & Weenink, 2011) using Lennes' Praat script (2003). (See the section, *Location of analyses within sentences*, above).

Results

All acoustic values were converted to Z-scores using the means and SDs of that measure for each speaker. All analyses were conducted using SPSS 24. We acknowledge that there is an imbalance between statements and *wh-* questions in the current corpus which mirrors the imbalance between the sentence types in naturalistic infant-directed speech (e.g., Newport, 1977; T. Wang, personal communication, July 20, 2020). To address this imbalance, we report both the overall correct classification rate for statements and *wh-* questions given by the binary logistic regression and how this compares to chance rates of classification for each analysis below.

Analysis 1: Wh-*Questions vs Statements when both start with two closed-class words*

We ran a binary logistic regression with Utterance (Wh_cc = *wh-* question starting with two closed-class words, and S_cc = statement starting with two closed-class words) as the classification variable and the six acoustic measurements (Duration, Mean F0, Max F0, Min F0, F0 range, and Intensity) in each word (w1 = word 1 and w2 = word 2) as the predictor variables. That means that the test is classifying the data into *wh-* questions and statements using the 12 predictors to assess whether the two utterances differ and which of the acoustic properties differentiate them. First, to find which of the 12 predictors contributed significantly to the classification of the data into statements and *wh-* questions, each predictor was entered in the model in a separate block (forward stepwise procedure). If the addition of each new predictor did not result in a statistically significant change in the model, then that predictor was removed from the final analysis (see Appendix B). Next, we ran a binary logistic regression with all the predictors found to be significant in the previous test, in one block. We found that the two utterance types differed in Duration, Mean F0, and Intensity of word 1, as well as in Duration and Intensity of word 2 (see Table 2). *Wh-* questions have shorter duration and higher Mean F0 and Intensity than statements (see Figure 1).

Overall, we found that there were differences in the acoustic properties at the beginning of the two utterances. That means that there are potential cues for infants to use when learning their language. In addition to the significance of the model and of the individual predictors, we can also assess the distinguishability of the two utterance categories with the classification results that are part of the binary logistic regression output. The binary logistic regression test initially classifies the data into the two categories tested (Wh_cc vs S_cc in this case) based only on their frequency. From this baseline model, we get the chance level of the classification. (68.4% in this case) Next, the

Table 2. Results from Binary Logistic Regression Comparing Wh_cc and S_cc.

	β (SE)	Wald's χ^2 (df = 1)	odds ratio (e^{β})	95% CI for odds ratio	
				Lower	Upper
Intercept	1.167 (0.083)*	196.994	3.214		
Duration – w1	0.353 (0.106)*	11.059	1.424	1.156	1.754
Mean F0 – w1	−0.417 (0.083)*	25.501	0.659	0.561	0.775
F0 Range – w1	0.109 (0.066)	2.750	1.115	0.980	1.268
Intensity – w1	−0.546 (0.084)*	42.562	0.579	0.492	0.683
Duration – w2	0.595 (0.112)*	28.170	1.814	1.456	2.260
Max F0 – w2	0.122 (0.093)	1.750	1.130	0.943	1.355
Intensity – w2	0.216 (0.076)*	8.179	1.241	1.070	1.440

Note. This table compares *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with two closed-class words (S_cc) for the first (w1) and second words (w2). The table lists all the predictors included in the model. Acoustic properties not included in the table were removed from the model based on the initial binary logistic regression. *$p < .05$; $R^2 = 0.104$ (Cox-Snell), 0.146 (Nagelkerke); -2 Log likelihood = 1536.607, Model χ^2 (7) = 148.318, $p < .05$, Hosmer-Lemeshow χ^2 (8) = 13.633, $p > .05$, overall correct classification = 70.7% (chance = 68.4%).

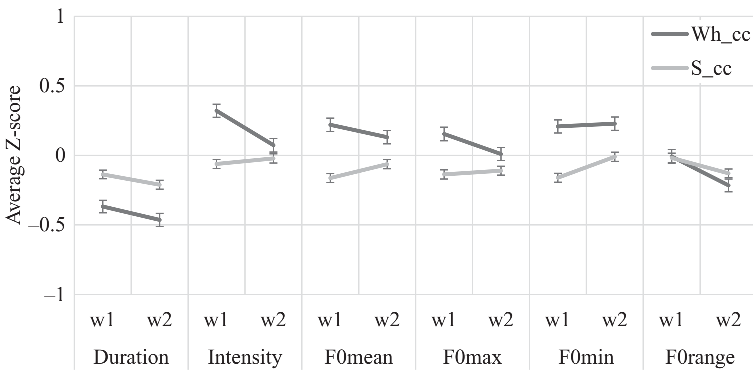

Figure 1. Z-scores per Measurement and Word Position for Wh_cc and S_cc.
Note. This graph compares measurements of the first (w1) and second words (w2) in *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with two closed-class words (S_cc). Error bars show the standard error of the mean.

test classifies the data into the two categories using all the predictors (for our case, the acoustic properties in Table 2). From this main model, we can see how much of the data were correctly classified in each category using our predictors.

In our test, even though there were statistically significant differences between the two utterances for some acoustic properties, those differences could only correctly classify 70.7% of the utterances into *wh-* questions and statements.[6] This means that these acoustic cues could help infants identify the correct utterance type a little above chance.

[6]Note that the chance level is 68.4% since there are more statements ($N_{S\_cc} = 923$) than *wh*-questions ($N_{Wh\_cc} = 427$).

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

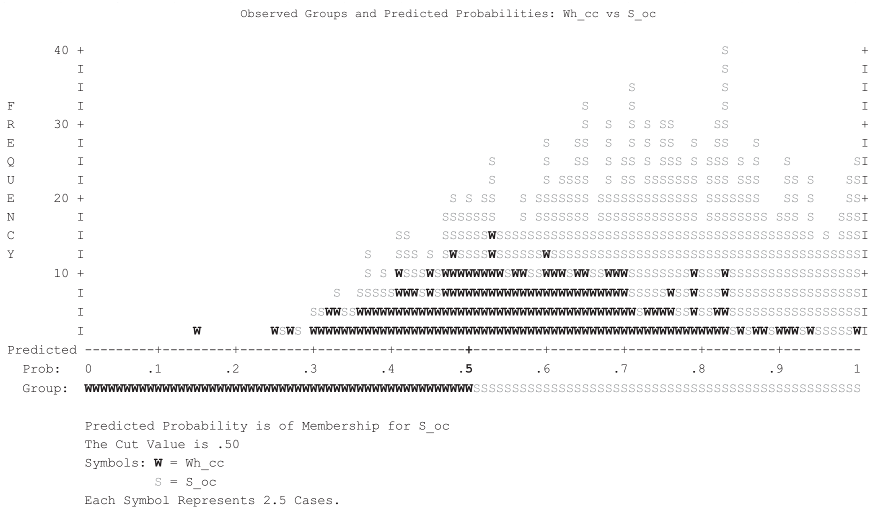

Figure 2. Histogram of the Predicted Probabilities of an Utterance Being a Statement (S_cc).
Note. The *wh*-question category is in the probability range of 0-0.5 and the statement category in the 0.5-1 range. That means that *wh*- questions that start with two closed-class words (Wh_cc) data in the 0.5-1 range are misclassified as statements, while statements that start with two closed-class words (S_cc) in the 0-0.5 range are misclassified as *wh*-questions

If we further break down the success rate, we see that while statements were correctly classified at 92.6%, *wh*- questions were only correctly classified at 23.4% (see Figure 2). This means that the acoustic differences between the two utterances was not enough to overcome the bias towards statements (i.e., the high frequency category), even though they did increase the successful classification of the utterances by 2.3%.

Since statements may start with different combinations of open-class and closed-class words, while the *wh*- questions we analyzed in the current corpus typically start with two closed-class words (with the exception of the 16 *wh*- questions mentioned previously that began with a closed-class word followed by an open-class word), there may be prosodic differences between the two sentence categories that are related to the presence and number of open-class words in utterance-initial positions. In our corpus, there were many statements that started with one or two open-class words, suggesting that the input to infants is much richer than a simple contrast between *wh*- questions and statements that start with two closed-class words. Even though these two utterance types do not differ much in the acoustic properties at the beginning of the utterance, it is possible that other types of statements differ more from *wh*- questions, and thus might be more useful for separating utterance types. To further explore this possibility, we ran three additional tests: we compared (a) *wh*- questions with statements starting with two open-class words, (b) *wh*- questions with statements starting with an open-class word followed by a closed-class word, and finally (c) *wh*-questions with statements starting with a closed-class word followed by an open-class word.

Analysis 2: Wh-*Questions vs. Statements starting with two open-class words*

Following the same method as in Analysis 1, to compare *wh*-questions (Wh_cc) with statements that start with two open-class words (S_oo), we first ran a binary logistic

Table 3. Results from Binary Logistic Regression Comparing Wh_cc and S_oo.

	β (SE)	Wald's χ^2 (df = 1)	odds ratio (e^{β})	95% CI for odds ratio	
				Lower	Upper
Intercept	−0.494 (0.116)*	18.303	0.610		
Duration – w1	0.116 (0.154)	0.562	1.122	0.830	1.518
Mean F0 – w1	−0.459 (0.132)*	12.131	0.632	0.488	0.818
Intensity – w1	−1.495 (0.155)*	92.639	0.224	0.165	0.304
Duration – w2	2.044 (0.213)*	92.295	7.722	5.089	11.717
Mean F0 – w2	−0.428 (0.282)	2.295	0.652	0.375	1.134
Min F0 – w2	0.785 (0.283)*	7.723	2.193	1.260	3.817
Intensity – w2	1.174 (0.157)*	56.087	3.236	2.380	4.401

Note. This table compares *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with two open-class words (S_oo) for the first (w1) and second words (w2)
The table includes all the predictors included in the model. Acoustic properties not included in the table were removed from the model based on the initial binary logistic regression.
*$p < .05$; $R^2 = 0.425$ (Cox-Snell), 0.579 (Nagelkerke); -2 Log likelihood = 532.146, Model χ^2 (7) = 381.175, $p < .05$, Hosmer-Lemeshow χ^2 (8) = 64.534, $p < .05$, overall correct classification = 84% (chance = 62.1%).

regression with each acoustic property in a separate block (see Appendix C) and then we ran another binary logistic regression with all the predictors that were significant in the first binary regression in one block. Duration of word 2, Mean F0 of word 1, Min F0 of word 1 and Intensity of words 1 and 2 were significant cues that differentiated *wh*-questions from statements (see Table 3). Figure 3 shows that *wh*-questions have shorter duration and higher F0 and intensity than statements.

Overall, we find that there were acoustic differences between the two utterance types. The classification results also support that there are sufficient acoustic cues that differentiate *wh-* questions from statements. The data were correctly classified 84% of the time (21.9% higher than chance),[7] with *wh-* questions having a success rate of 91.8% and statements of 71.3%. As Figure 4 also shows, the acoustic cues classify the utterances into the two types pretty well.

Analysis 3: Wh-*Questions vs. Statements starting with an open-class and a closed-class word*

Next, we compared the *wh-* questions (Wh_cc) to the statements that began with an open-class word followed by a closed-class word (S_oc). In the binary logistic regression with all the significant predictors in a single block (see Appendix D for the initial analysis), we found that Duration and Intensity of words 1 and 2 were significant cues (Table 4). Figure 5 shows that *wh-* questions have shorter duration and higher intensity than statements.

[7]Note that the chance level is 62.1% since there are more *wh-* questions ($N_{Wh\_oo} = 427$) than statements ($N_{S\_oo} = 261$).

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

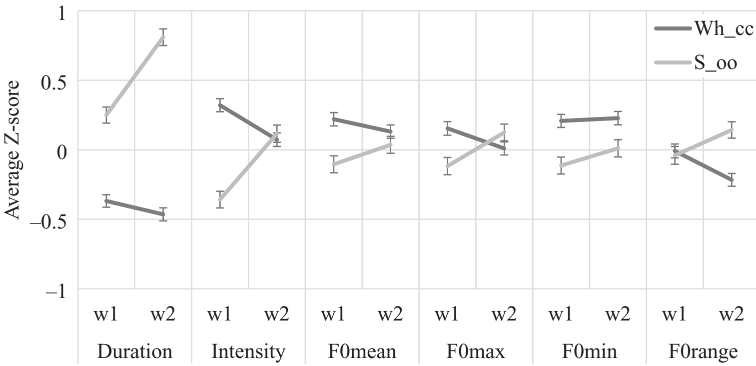

Figure 3. Z-Scores per Measurement and Word Position for Wh_cc and S_oo.
Note. This graph compares measurements of the first (w1) and second words (w2) in *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with two open-class words (S_oo). Error bars show the standard error of the mean.

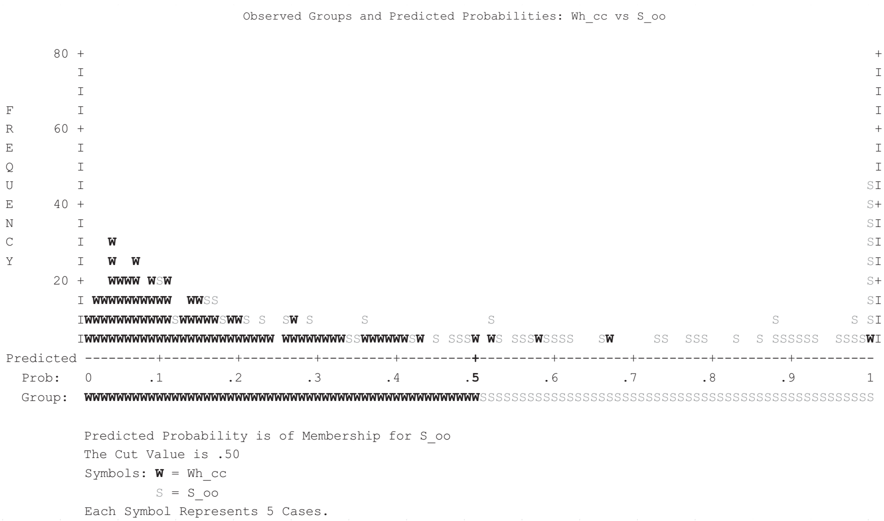

Figure 4. Histogram of the Predicted Probabilities of an Utterance Being a Statement (S_oo).
Note. The *wh*-question category is in the probability range of 0-0.5 and the statement category in the 0.5-1 range. That means that *wh-* questions that start with two closed-class words (Wh_cc) data in the 0.5-1 range are misclassified as statements, while statements that start with two open-class words (S_oo) in the 0-0.5 range are misclassified as *wh-* questions.

Overall, our results indicate that there are acoustic cues that distinguish *wh-* questions from statements, but looking at the classification, we see that the success rate is low. Only 72.1% of the data were correctly classified into *wh-* questions and statements (3.5% higher than chance).[8] The statements were classified 89.2% correctly, while the *wh-* questions only 34.9% (also in Figure 6).

[8]Note that the chance level is 68.6% since there are more statements (N_{S_oc} = 931) than *wh*-questions (N_{Wh_cc} = 427).

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

Table 4. Results from Binary Logistic Regression Comparing Wh_cc and S_oc.

	β (SE)	Wald's χ^2 (df = 1)	odds ratio (e^{β})	95% CI for odds ratio	
				Lower	Upper
Intercept	1.144 (0.081)*	200.643	3.141		
Duration – w1	0.49 (0.097)*	25.497	1.632	1.349	1.974
Max F0 – w1	0.009 (0.085)	0.011	1.009	0.854	1.192
Intensity – w1	−0.807 (0.089)*	82.332	0.446	0.375	0.531
Duration – w2	0.815 (0.114)*	50.924	2.259	1.806	2.825
Max F0 – w2	0.167 (0.089)	3.488	1.181	0.992	1.407
Intensity – w2	0.499 (0.083)*	36.370	1.647	1.401	1.938

Note. This table compares *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with an open-class word followed by a closed-class word (S_oc) for the first (w1) and second words (w2). The table includes all the predictors included in the model. Acoustic properties not included in the table were removed from the model based on the initial binary logistic regression.
*$p < .05$; $R^2 = 0.159$ (Cox-Snell), 0.223 (Nagelkerke); -2 Log likelihood = 1456.115, Model χ^2 (6) = 234.871, $p < .05$, Hosmer-Lemeshow χ^2 (8) = 16.86, $p < .05$, overall correct classification = 72.1% (chance = 68.6%).

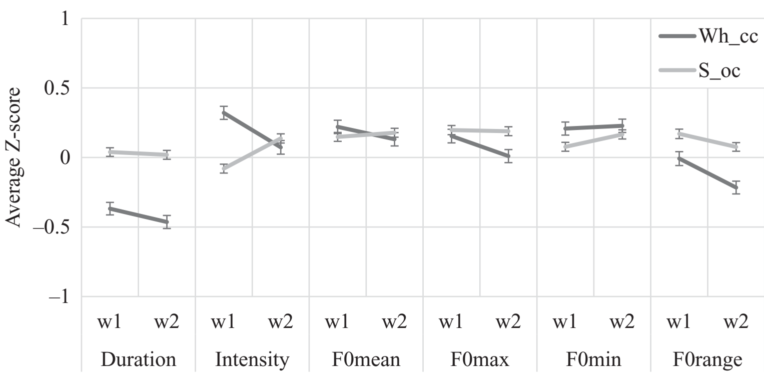

Figure 5. Z-Scores per Measurement and Word Position for Wh_cc and S_oc.
Note. This graph compares measurements of the first (w1) and second words (w2) in *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with an open-class word followed by a closed-class word (S_oc). Error bars show the standard error of the mean.

Analysis 4: Wh*-Questions vs. Statements starting with a closed-class and an open-class word*

Finally, we compared *wh-* questions (Wh_cc) to statements that start with a closed-class word followed by an open-class word (S_co). In the binary logistic regression with all the significant predictors in a single block (see Appendix E for the initial analysis) we found that Duration and Intensity of words 1 and 2, and Mean F0 of word 1 were significant predictors in the classification of the data into *wh-* questions and statements (Table 5). In Figure 7, we see that *wh-* questions have shorter duration and higher F0 and Intensity.

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

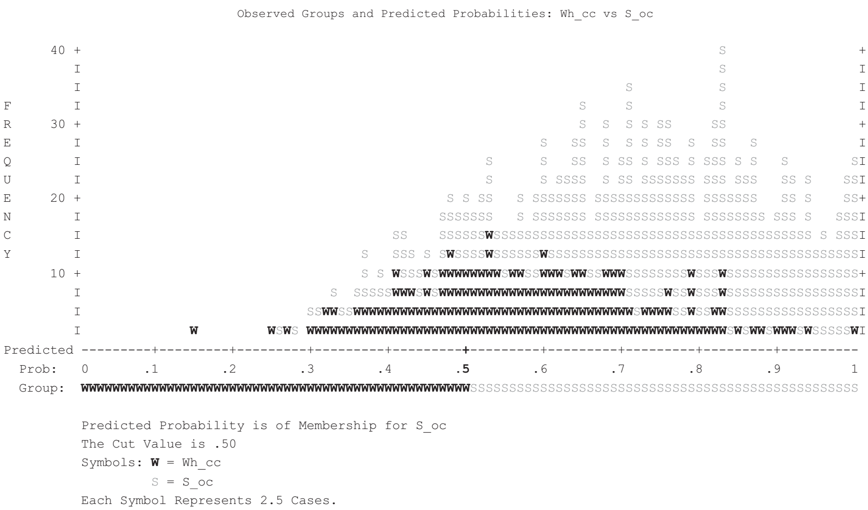

Figure 6. Histogram of the Predicted Probabilities of an Utterance Being a Statement (S_oc).
Note. The *wh-* question category is in the probability range of 0-0.5 and the statement category in the 0.5-1 range. That means that *wh-* questions that start with two closed-class words (Wh_cc) data in the 0.5-1 range are misclassified as statements, while statements that start with an open-class word followed by a closed-class word (S_oc) in the 0-0.5 range are misclassified as *wh*-questions.

Table 5. Results from Binary Logistic Regression Comparing Wh_cc and S_co.

	β (SE)	Wald's χ^2 (df = 1)	odds ratio (e^{β})	95% CI for odds ratio	
				Lower	Upper
Intercept	0.841 (0.088)*	90.452	2.318		
Duration – w1	−0.182 (0.086)*	4.501	0.833	0.704	0.986
Mean F0 – w1	−0.503 (0.101)*	24.946	0.605	0.497	0.737
F0 Range – w1	0.002 (0.09)	0.000	1.002	0.839	1.196
Intensity – w1	−0.926 (0.115)*	65.021	0.396	0.316	0.496
Duration – w2	2.32 (0.162)*	206.246	10.178	7.416	13.970
Max F0 – w2	0.138 (0.106)	1.717	1.148	0.934	1.413
Intensity – w2	0.695 (0.105)*	44.067	2.005	1.632	2.461

Note. This table compares *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with a closed-class word followed by an open-class word (S_co) for the first (w1) and second words (w2). The table includes all the predictors included in the model. Acoustic properties not included in the table were removed from the model based on the initial binary logistic regression.
*$p < .05$; $R^2 = 0.358$ (Cox-Snell), 0.492 (Nagelkerke); -2 Log likelihood = 1030.84, Model χ^2 (7) = 531.52, $p < .05$, Hosmer-Lemeshow χ^2 (8) = 91.645, $p < .05$, overall correct classification = 81.3% (chance = 64.4%).

Overall, we found that there were acoustic cues at the beginning of an utterance to distinguish *wh-* questions from statements that began with a closed-class word followed by an open-class word. The classification results also supported this. We found that 81.3%

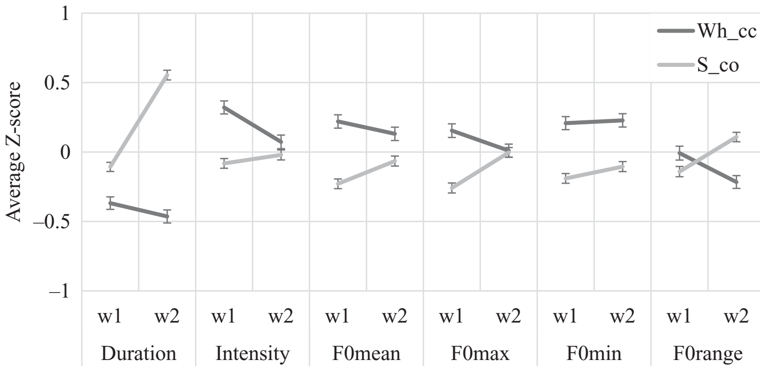

Figure 7. Z-Scores per Measurement and Word Position for Wh_cc and S_co.
Note. This graph compares measurements of the first (w1) and second words (w2) in *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with a closed-class word followed by an open-class word (S_co). Error bars show the standard error of the mean.

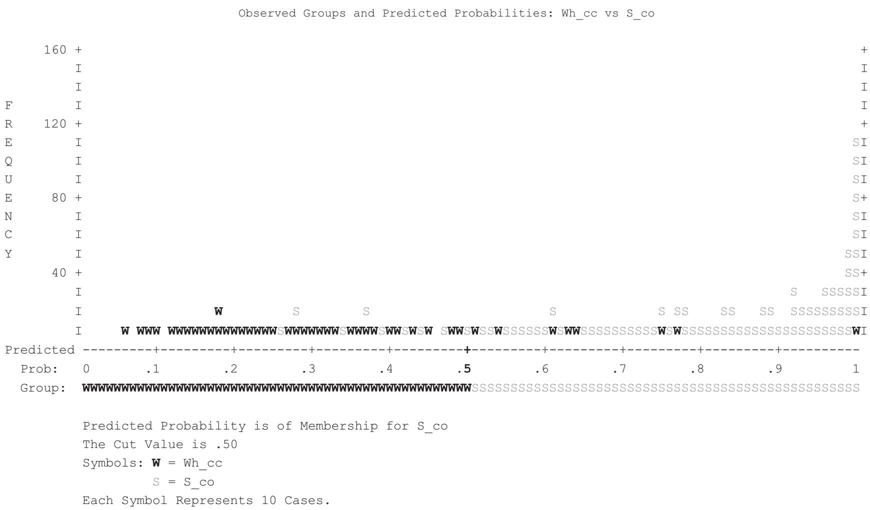

Figure 8. Histogram of the Predicted Probabilities of an Utterance Being a Statement (S_co).
Note. The *wh*-question category is in the probability range of 0-0.5 and the statement category in the 0.5-1 range. That means that *wh-* questions that start with two closed-class words (Wh_cc) data in the 0.5-1 range are misclassified as statements, while statements that start with a closed-class word followed by an open-class word (S_co) in the 0-0.5 range are misclassified as *wh*-questions.

of the data were correctly classified into the two utterances (16.9% higher than chance).[9] The *wh-* questions were correctly classified 70.3% of the time and the statements 87.3% (also in Figure 8).

[9]Note that the chance level is 64.4% since there are more statements (N_{S_co} = 773) than *wh*-questions (N_{Wh_cc} = 427).

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

Table 6. Summary of Results From All Comparisons

Wh_cc vs	-2LL	Correct Classification				R^2	
		Overall	(chance)	Wh-question	Statement	Cox-Snell	Nagelkerke
S_cc	1536.607	70.7%	(68.4%)	23.4%	92.6%	0.104	0.146
S_oo	532.146	84.0%	(62.1%)	91.8%	71.3%	0.425	0.579
S_oc	1456.115	72.1%	(68.6%)	34.9%	89.2%	0.159	0.223
S_co	1030.840	81.3%	(64.4%)	70.3%	87.3%	0.358	0.492

Note. This table summarizes comparisons between *wh-* questions that start with two closed-class words (Wh_cc) and statements that start with two closed-class words (S_cc), statements that start with two open-class words (S_oo), statements that start with an open-class word followed by a closed-class word (S_oc), and statement that start with a closed-class word followed by an open-class word (S_co).

Summary

We compared the results from the comparisons of *wh-* questions to the four types of statements: (1) beginning with two closed-class words (S_cc), (2) beginning with two open-class words (S_oo), (3) beginning with an open-class word followed by a closed-class word (S_oc), and (4) beginning with a closed-class word followed by an open-class word (S_co). The results are summarized in Table 6.

The best model (i.e., the model that fits best to the data) was that of S_oo, which has the lowest deviance (-2LL) and the highest classification rate. That means that out of the four comparisons, the acoustic differences between *wh-* questions and statements distinguished best the questions from statements that begin with two open-class words. S_co also showed a good fit, indicating that *wh-* questions were also well distinguished from statements that began with a closed-class plus an open-class word. In contrast, S_cc and S_oc had the highest deviance and lower classification rates with the data showing a bias towards statements. That means that *wh-* questions were not well distinguished from statements when the statements began with two closed-class or an open-class plus a closed-class word. So, the statements that were more distinguishable from *wh-* questions were those that had an open-class word as their second word, while statements with a second closed-class word were very similar to *wh-* questions.

The differences between the utterances were mainly in duration and intensity.

Figure 9 summarizes all the acoustic properties for each utterance. The *wh-* questions began with shorter words than statements (not surprising, since the first word of a statement is often an open-class word, and might entail less vowel reduction or have more a complex onset/coda than *wh-* words), and especially for the statements with a second open-class word, the difference became even larger in the second word position (w2). On the other hand, *wh-* questions had a higher intensity than statements, especially in first word position (w1). With respect to the F0 properties (mean, max, min, range), in each comparison, a different property was significant with the exception of F0 range, which was never a significant factor. Overall, *wh-* questions started with higher F0 than statements, but the statements with an open-class plus a closed-class word were the most similar to the *wh-* questions.

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

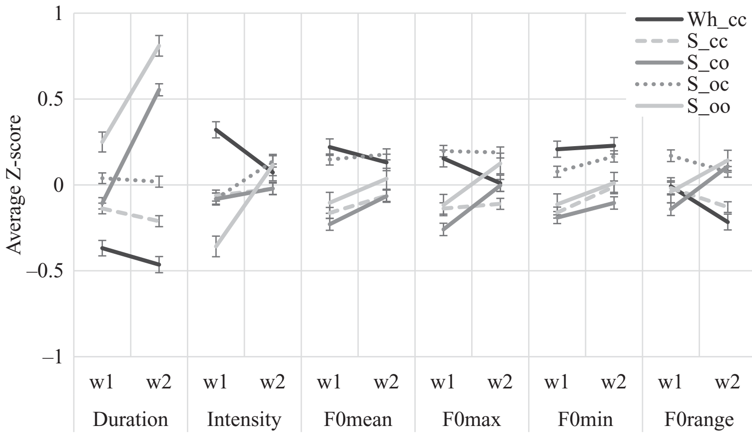

Figure 9. Acoustic Properties of Each Utterance Type by Word.
Note. This graph compares measurements of the first (w1) and second words (w2) in *wh-* questions that start with two closed-class words (Wh_cc), statements that start with two closed-class words (S_cc), statements that start with a closed-class word followed by an open-class word (S_co), statements that start with an open-class word followed by a closed-class word (S_oc), and statements that start with two open-class words (S_oo). Error bars show the standard error of the mean.

Discussion

The fact that the specific combination of prosodic cues needed for sentence-type classification varies at the sentence-level as well as the word level (whether sentences start with an open- or closed-class word) highlights the challenges in developing a comprehensive account of how infants could use prosody to make useful sentence-type discriminations. We hypothesized that there would be broad utterance-initial prosodic differences between *wh-* questions and statements in American English infant-directed speech depending on the first words (closed-class) that appear in *wh-*questions compared to the variety of first words found in statements (open- and closed-class) and our results supported this hypothesis. However, although the results show that a variety of prosodic cues provide information that infants could use to broadly distinguish between sentence types, few studies have evaluated how useful these combinations of cues are for actually categorizing individual utterances. Previous studies (Chiang et al., 2018; Geffen & Mintz, 2017) found that models of English IDS rarely exceeded 50-60% for correctly categorizing statements and *wh-* questions, whether evaluating information at the beginning (Chiang et al., 2018) or the end of utterances (Geffen & Mintz, 2017). However, both previous studies utilized a small corpus (approximately 300 utterances) divided roughly equally between three sentence types: statements, yes/no questions and *wh-* questions. A larger sample of ~800 Dutch utterances were automatically classified with high levels of accuracy into broad categories of statements and questions (82% and 90% accuracy respectively) though question sub-types (yes/no and *wh-* questions) had lower accuracy (53-75%; van Heuven et al., 1997). We wanted to determine whether providing more experience (in this case a greater number of phrases to learn from) would improve the model for English sentence-type discrimination. We chose to focus on the distinction between statements and *wh-* questions given previous findings about the similarities of these two sentence types at the end of utterances. We wanted to see if prosodic information was available at the beginning of the utterances that would provide cues for infants to make initial sentence-type distinctions. We also added additional sets of comparisons, comparing *wh-* questions to four different types of statement phrases to see if prosodic information would provide sufficient information to correctly categorize and distinguish *wh-* questions from the different types of statements. However, the decision to look at statements with different phrase structures (e.g., statements that begin with two closed-class words versus two open-class words) does raise questions about the generalizability of the results to all statements. It may be that, as with many other categories, there is more variability within one type of phrasal structure (e.g., statements that begin with two closed-class words) than there are between sentence types (e.g., *wh-* questions that begin with two closed-class words versus statements that begin with two closed-class words).

We predicted that prosodic cues (e.g., Mean F0, Intensity) and word category (open vs. closed-class) would serve to correctly classify utterances as either statements or *wh-* questions in the models discussed in the Results section. Our results partly supported this hypothesis. Collectively, the classification results demonstrated several patterns. First, prediction accuracy for sentence-type increases with sample size, a pattern we see through all four analyses (with the exception of Analysis 2, the statement category generally had a higher sample size). This is not surprising given that there is bias in the test itself towards the category with the highest frequency. Within our corpus, the statements had a much higher frequency than *wh-* questions. Even when the sample contained more questions than statements (as in Analysis 2), classification accuracy did not fall below 70% for statements, while there was greater variability for *wh-* questions (23.4%-91.8%). This suggests that acoustic cues are useful for classifying both sentence types, although these models are better at classifying statements than questions. Future research should examine how the model would perform if the corpus had a different ratio of statements and questions. For example, in the corpus in Newport (1977), *wh-* questions accounted for approximately 21% of the input directed toward children, while they only accounted for approximately 15% of the current corpus. The current study suggests that an imbalance of statements and (*wh-*) questions may be the norm in infant-directed speech, with fewer questions in the daily input (in line with previous studies, e.g., Newport, 1977; T. Wang, personal communication, July 20, 2020). This raises the question of whether an increase of *wh-* questions in the input lead to greater classification accuracy for the model overall or only for other instances of *wh-* questions? More importantly, how much input (particularly for different question types) do infants require to increase their classification accuracy? Further research is needed to determine whether we can change the bias in the test itself when we add predictors, although our initial results suggest not. This also raises questions about what this kind of input means for the learning process.

Additionally, the word categories of the first two words in the sentence types, and the second word in particular, seem to impact categorization accuracy. When the second word in the phrase was a closed-class word (Analyses 1 and 3), there was a large difference in classification accuracy between statements and questions (69.2% and 54.3% respectively), whereas when the second word was an open-class word (Analyses 2 and 4), the difference was smaller (20.5% and 17% respectively). There was a similar difference in overall accuracy with phrases that had an open-class word as the second word; these demonstrated a greater difference from chance (21.9% and 16.9% greater than chance respectively) compared to phrases that had a closed-class word as the second word (2.3% and 3.5% greater than chance respectively). The average overall classification accuracy across all four models was 11.2% different from chance. This is in line with results from Geffen and Mintz (2017), which found that statements and questions were correctly identified 51% of the time and Chiang et al. (2018) who found a classification rate of 59% (chance was approximately 50% for both). While prosodic information is available to distinguish between *wh-* questions and different types of statements, the current models varied in overall classification accuracy. The question remains how much of this is due to the different phrasal structures of the statements in the current corpus analysis (e.g., statements that begin with two closed-class words versus two open-class words). Thus, prosody could be a useful cue, but not likely the driving force behind infants' initial sentence-type discrimination, something that should be evaluated in future infant perception studies. It is unclear whether 2.3% and 3.5% higher than chance would be enough for infants (or human categorization in general) to create those categories, though it seems unlikely.

Previous studies demonstrated that prosodic information (especially pitch) is available at the beginning (Chiang et al., 2018) but not the end of utterances (Geffen & Mintz, 2017) to distinguish between statements and *wh-* questions. The current study provides support for Chiang et al. (2018), finding prosodic differences between sentence types in Mean F0, Duration and Intensity. These results are also consistent with the Edge hypothesis (Seidl & Johnson, 2006), which suggests that infants pay special attention to initial words. For example, children can learn auxiliary verbs when they are utterance initial (yes/no questions) but not utterance medial (*wh-* questions; Newport, Gleitman & Gleitman, 1977). Our results provided partial support for the Edge hypothesis, suggesting that initial words are relevant for making sentence-type distinctions. There are acoustic properties at the beginning of utterances that could distinguish *wh-* questions from statements, but the degree of distinguishability and the usefulness of the cues to the learning infant (i.e., the classifier) depends on the types of words used (open-class vs. closed-class) as well as the position (the category of the SECOND word was more important for classification accuracy than the category of the first word). Further research is needed to tease apart the impact of word category and position.

Results also demonstrated consistent differences in duration across the four analyses, such that *wh-* questions were always shorter than statements. As we discussed in the introduction, duration is often considered to be a secondary prosodic characteristic, varying along with pitch, but not necessarily the characteristic driving discrimination. For example, Geffen and Mintz (2017) found no durational differences between the ends of statements and *wh-* questions, although questions generally have shorter (overall) duration than statements (a pattern we saw in the utterance-initial words in the current corpus analysis). Nonetheless, in the current study duration was a significant predictor of classification accuracy in all four of the models, consistently showing significant differences for word 2 and, apart from Analysis 2, differences in word 1 as well.

The current results raise the question of whether infants are making distinctions at the sentence (statements vs. questions) or word level (closed-class vs. open-class). For example, all the *wh-* questions included in the current corpus began with two closed-class words (only 16 *wh-* questions out of 443 did not follow this pattern). In contrast, the current corpus analysis examined four different types of statements that differed in the category of their first and second word (cc, co, oo, oc). While looking at different types of statements provided an opportunity to look at word level distinctions, it likely also had the side effect of decreasing effect sizes in differences between statements and *wh-* questions. In future analyses, it might be more informative to contrast different types of phrasal structures (e.g., combining statements and *wh-* questions that begin with two closed-class words and comparing them to sentences that begin with two open-class words). Regardless of the type of distinction, these results suggest that there is a relationship between prosodic and syntactic information in an utterance-initial position that infants may be able to leverage to make sentence-type distinctions, though likely not at the 9- to 10-month age of the infants included in the current corpus analysis.

One limitation of the current study is that we did not take context into account when examining statements and questions. Given that there is not much work in this field, it was important to establish whether prosodic information is available at the beginning (Chiang et al., 2018) as well as the end of statements and questions (Geffen & Mintz, 2017). However, sentences do not occur in a vacuum. Prosody changes within individual sentences as well as over the course of a conversation. Future research should examine whether there are discourse level changes in prosodic information that highlight the beginning of utterances as an important source of information for classifying and distinguishing between statements and questions.

Another potential limitation has to do with our selection criteria. Coders were instructed not to include any *wh-* questions where they did not feel that the mother's intention was to ask a question (one of the exclusionary criteria in Appendix A). Our selection criteria were based on a combination of prosodic and syntactic features (e.g., utterance initial *wh-* word followed in many cases by an auxiliary verb or do-support) but this could have introduced an element of selection bias. Future studies should evaluate this potential bias by looking at all utterances that begin with *wh-* words, whether they have question or statement word order and/or prosodic cues (e.g., final flat or falling intonation).

The current study found (more pronounced) prosodic differences between *wh-* questions and different types of statements based on whether the first two words were open-class or closed-class words. While this result was unexpected, it makes sense given the prosodic and syntactic differences between open-class and closed-class words. This raises a question that was also posed by Chiang et al. (2018): Are there similar types of distinctions between different types of *wh-* questions? We know that infants' comprehension and production of different types of *wh-* questions proceeds in a fairly consistent order (Rowland et al., 2003). However, the question remains of whether this is due entirely to cognitive understanding of what the *wh-* question is asking or whether there is an additional prosodic element that makes it easier to understand certain questions sooner than others. Prosody can help disambiguate between different sentence forms, but syntactic complexity takes longer to acquire as shown by *wh-* question studies which consistently demonstrate that object *wh-*questions, which have greater syntactic complexity, are acquired later. It would be interesting to see whether once infants or toddlers have acquired these different types of *wh-* questions, they weight prosodic cues differently for each type. It is also not surprising that infants acquire subject *wh-* questions first considering that words like "what" are more common in their daily input than "how" (Rowland et al., 2003). Therefore, it makes sense that prosody may become more variable as syntactic complexity increases. If this is true in the early stages, it makes sense that we would see the consistent order of acquisition for *wh-* words and *wh-* questions that has been demonstrated by previous studies (Bloom, Merkin & Wooten, 1982; Rowland et al., 2003).

Conclusion

In summary, the current experiment provides preliminary evidence that prosodic information could be useful for distinguishing statements from *wh-* questions in infant-directed speech, although it remains to be seen whether infants can use these cues to make sentence-type distinctions by the time they are 1;0 year. Making this distinction could perhaps provide a foundation for distinguishing *wh-* questions from statements on utterance-initial distributional grounds, as we have found that those sentence types are prosodically similar at the end of utterances in infant-directed and adult-directed speech. Our results suggest that differing combinations of prosodic and segmental cues are available for infants to make broad distinctions between statements and questions. These differences could provide an important foundation for acquiring syntactic knowledge. Future research is needed to further evaluate infants' sensitivity to prosodic and segmental differences between statements and questions, as well as between categories of words, specifically auxiliary verbs (e.g., can, do) and *wh-* words.

Acknowledgements. This research was supported by funding from Natural Sciences and Engineering Research Council of Canada [NSERC grant number 03797-2018, awarded to S. Curtin]. Funding agencies had no role in study design, data collection or analysis, or in the writing/submission of this paper. We would like to thank the members of the University of Calgary Speech Development Lab for their help with this research, especially our volunteer coders.

Competing interests. The authors declare none.

References

Bartels, C. (1999). *The intonation of English statements and questions: A compositional interpretation*. New York & London, Garland Publishing Inc.

Bergelson, E., **Casillas, M.**, **Soderstrom, M.**, **Seidl, A.**, **Warlaumont, A. S.**, & **Amatuni, A.** (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, **22**(1), e12724.

Best, C., **Levitt, A.**, & **McRoberts, G.** (1991). Examination of language-specific influences in infants' discrimination of prosodic categories. In *Proceedings of the 12th International Congress of Phonetic Sciences* (Vol. 4, pp. 162–165).

Bloom, L., **Merkin, S.**, & **Wooten, J.** (1982). *Wh*-questions: Linguistic factors that contribute to the sequence of acquisition. *Child Development*, **53**, 1084–1092.

Boersma, P., & **Weenink, D.** (2011). Praat: Doing phonetics by computer (version 5.2.23) [Computer program]. Available from www.praat.org.

Brent, M., & **Siskind, J.** (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, **81**, 31–44.

Chiang, C., **Geffen, S.**, & **Mintz, T. H.** (2018). Distinguishing questions and statements using sentence-initial prosodic cues. In Bertolini, A. B., & Kaplan, M.J. (Eds.) *Proceedings of the 42nd Boston University Conference on Language Development (BUCLD), Boston University* (pp. 153–158). Somerville, MA: Cascadilla Press.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Face, T. (2007). The role of intonation cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de fonética experimental*, **16**, 185–225.

Falé, I., & **Faria, I. H.** (2006). Time course of intonation processing in European Portuguese: A gating paradigm approach. *IV Jornadas en Tecnologia del Habla*, 265–270.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (5–29). Norwood, NJ: Ablex.

Fernald, A., & **Kuhl, P.** (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 279–93.

Fernald, A., & **Morikawa, H.** (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, **64**(3), 637–656.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, **27**(4), 765–768.

Geffen, S., & **Mintz, T. H.** (2017). Prosodic differences between declaratives and interrogatives in infant-directed speech. *Journal of Child Language*, **44**(4), 968–994. https://doi.org/10.1017/S0305000916000349

Gerken, L., **Jusczyk, P.**, & **Mandel, D.** (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, **51**, 237–265.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters, (Eds.), *Regional variation in intonation (*9-31*).* Tubingen: Niemeyer.

Gussenhoven, C., & **Chen, A.** (2000). Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing (ICSLP 2000)* (pp. 91–94).

Hay, J., & **Saffran, J.** (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, **17**, 610–641.

Hedberg, N., & **Sosa, J.** (2002). The prosody of questions in natural discourse. In *Speech Prosody 2002, International Conference.*

Hedberg, N., **Sosa, J.**, & **Fadden, L.** (2004). Meanings and configurations of questions in English. In *Proceedings of Speech Prosody 2004,* Nara, Japan, (pp. 309–312).

Hedberg, N., **Sosa, J.**, **Görgülü, E.**, & **Mameni, M.** (2010). The prosody and meaning of wh-questions in American English. In *Speech Prosody 2010-Fifth International Conference.*

Hirsh-Pasek, K., **Kemler Nelson, D.**, **Jusczyk, P.**, **Cassidy, K.**, **Druss, B.**, & **Kennedy, L.** (1987). Clauses are perceptual units for young infants. *Cognition*, **26**, 269–286.

Hochmann, J., **Endress, A.**, & **Mehler, J.** (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, **115**(3), 444–457.

Höhle, B., & **Weissenborn, J.** (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, **6**(2), 122–127.

Jarman, E., & **Cruttenden, A.** (1976). Belfast intonation and the myth of the fall. *Journal of the International Phonetic Association*, **6**, 4–12.

Jusczyk, P., **Hirsh-Pasek, K.**, **Kemler Nelson, D.**, **Kennedy, L.**, **Woodward, A.**, & **Piwoz, J.** (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, **24**, 252 – 293.

Kemler Nelson, D., **Hirsh-Pasek, K.**, **Jusczyk, P.**, & **Cassidy, K.** (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language* **16**, 55–68.

Ladd, D. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, P., & **Johnson, K.** (2010). *A course in phonetics* (6^{th} ed). Stamford, CT: Cengage.

Lennes, M. (2003). collect_formant_data_from_files.praat. https://lennes.github.io/spect/

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

MacKenzie, H., **Curtin, S.**, & **Graham, S.** (2012). Class matters: 12-month-olds' word–object associations privilege content over function words. *Developmental Science*, **15**(6), 753–761. https://doi.org/10.1111/j.1467-7687.2012.01166.x

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maxwell, O., & **Fletcher, J.** (2013). The prosody of WH-questions in Bengali and Kannada English. Paper presented at the Australian Linguistic Society conference, Melbourne, Australia.

McRoberts, G. W., **McDonough, C.**, & **Lakusta, L.** (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy*, **14**(2), 162–194.

Mehler, J., **Jusczyk, P.**, **Lambertz, G.**, **Halsted, N.**, **Bertoncini, J.**, & **Amiel-Tison, C.** (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 143–178.

Monaghan, P., **Christiansen, M.**, & **Chater, N.** (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, **55**(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Moon, C., **Cooper, R.**, & **Fifer, W.** (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, **16**, 495–500.

Morgan, J., **Shi, R.**, & **Allopenna, P.** (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J.L Morgan & K. Demuth (Eds.)., *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 263–283. Lawrence Erlbaum Associates.

Nazzi, T., **Bertoncini, J.**, & **Mehler, J.** (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology*, **24**, 756–766.

Nespor, M., **Shukla, M.**, & **Mehler, J.** (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewan, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology: Vol. 2* (1147-1159). Malden, MA and Oxford: Wiley-Blackwell.

Newport, E. (1977). Motherese: The speech of mothers to young children. In N. J. Castellan, D. B. Pisoni & G. R. Potts (Eds.), *Cognitive theory: Vol. 2*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Newport, E., **Gleitman, H.**, & **Gleitman, L.** (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children (pp. 109-149).* Cambridge, UK: Cambridge University Press.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics*, **7**, 119–145.

Patel, R., & **Grigos, M.** (2006). Acoustic characterization of the question-statement contrast in 4, 7, and 11 year-old children. *Speech Communication*, **48**, 1308–1318.

Patience, M., **Marasco, O.**, **Colanton, L.**, **Klassen, G.**, **Radu, M.**, & **Tararova, O.** (2018). Initial pitch cues in English sentence types. In *Proceedings of 9th International Conference on Speech Prosody 2018* (pp. 463–467).

Petrone, C., & **Niebuhr, O.** (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, **57**(1), 108–146.

Rowland, C. F., **Pine, J. M.**, **Lieven, E. V.**, & **Theakston, A. L.** (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language*, **30**(3), 609–635.

Săfářŏvá, M., & **Swerts, M.** (2004). On recognition of declarative questions in English. In *Proceedings of Speech Prosody 2004,* Nara, Japan.

Saindon, M. R., **Trehub, S. E.**, **Schellenberg, E. G.**, & **van Lieshout, P. H.** (2017). When is a question a question for children and adults? *Language Learning and Development*, **13**(3), 274–285.

Seidl, A., **Hollich, G.**, & **Jusczyk, P. W.** (2003). Early understanding of subject and object wh-questions. *Infancy*, **4**(3), 423–436.

Seidl, A., & **Johnson, E.** (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, **9**, 565–573.

Shady, M. E. (1996). *Infants' sensitivity to function morphemes (Unpublished doctoral dissertation)*. State University of New York at Buffalo, Buffalo, NY.

Shi, R., **Cutler, A.**, **Werker, J.**, & **Cruickshank, M.** (2006a). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America*, **119**(6), EL61-EL67. https://doi.org/10.1121/1.2198947

Shi, R., & **Lepage, M.** (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, **11**(3), 407–413. https://doi.org/10.1111/j.1467-7687.2008.00685.x

Shi, R., **Marquis, A.**, & **Gauthier, B.** (2006b). Segmentation and representation of function words in preverbal French-learning infants. In *Proceedings of the 30th Annual Boston University Conference on Language Development* (Vol. 2, pp. 549–560). Somerville, MA: Cascadilla.

Shi, R., **Morgan, J.**, & **Allopenna, P.** (1999). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language* **25**(1), 169–201. https://doi.org/10.1017/S0305000097003395

Shi, R., & **Werker, J. F.** (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, **12** (1), 70–75.

Shi, R., **Werker, J. F.**, & **Morgan, J. L.** (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, **72**(2), B11–B21. https://doi.org/10.1016/S0010-0277(99)00047-5

Slobin, D., & **Bever, T.** (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, **12**, 229–265.

Soderstrom, M., **Blossom, M.**, **Foygel, R.**, & **Morgan, J. L.** (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, **35**(4), 869–902.

Soderstrom, M., **Ko, E.**, & **Nevzorova, U.** (2011). It's a question? Infants attend differently to polar questions and declaratives. *Infant Behavior and Development*, **34**, 107–110.

Stern, D. N., **Spieker, S.**, **Barnett, R. K.**, & **MacKain, K.** (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, **10**(1), 1–15.

van Heuven, V., & **Haan, J.** (2000). Phonetic correlates of statement versus question intonation in Dutch. In *Intonation* (pp. 119–143). Springer, Dordrecht.

van Heuven, V. J. V., **Haan, J.**, & **Pacilly, J. J.** (1997). Automatic recognition of sentence type from prosody in Dutch. In *Fifth European Conference on Speech Communication and Technology*.

van Heuven, V., & **van Zanten, E.** (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, **47**, 87–99.

Vion, M., & **Colas, A.** (2006). Pitch cues for the recognition of yes–no questions in French. *Journal of Psycholinguistic Research* **35**, 427–445.

Appendix A: Selection Criteria

Sentence Type	Inclusion criteria	Exclusion criteria
WH questions	Utterance-initial wh- word	Phrases that begin with discourse marker (e.g., yeah, oh, uh-oh, whoa)
	Mother's intent seems to be a question (as judged by native English speaker)	Phrases that begin with a *wh-* word but are not a question (e.g., what a good girl, how nice)
	May have question mark	
	Aux-inversion	
Statements	Canonical word order	Declarative question
		Partial initial word (e.g., stuttering)
		Initial dysfluency (uh, um) – can use sentence after dysfluency
		Phrases that end with "huh"
		Phrases that only consist of very common proper names equivalent to single words (e.g., "Chips Ahoy", "Mickey Mouse")
General criteria		One-word utterances (e.g., What? Stop.)
	Phrases with background noise at the end are okay.	Phrases with background noises obscuring wh- word or first word of statement (vocalization or background noise, for example: laughing, crying, blowing raspberries)
		Unintelligible phrases
		Phrases that include [sung, read] at the end*
	Phrases like "Say (Let's play)" or something similar are okay.	Phrases like "Say (whee, oops)". If it is just a sound after the word say, and nothing else, do not include it.
		Phrases that include [XXX] within the first couple words

*More routine, less naturalistic speech

Appendix B. Model results – Wh_cc vs S_cc

	-2LL	Data correctly classified	χ^2 (df = 1)	R^2 Cox & Snell	R^2 Nagelkerke
Block 0 (no predictors)	1684.9	68.4%			
Block 1 (Duration – w1)	1657.3	68.4%	27.649*	0.020	0.028
Block 2 (Mean F0 – w1)	1614.7	69.8%	42.609*	0.051	0.071
Block 3 (Max F0 – w1)	1613.7	69.3%	0.919	0.051	0.072
Block 4 (Min F0 – w1)	1612.6	69.4%	1.111	0.052	0.073
Block 5 (F0 Range – w1)	1605.2	69.5%	7.416*	0.057	0.080
Block 6 (Intensity – w1)	1568.0	69.3%	37.244*	0.083	0.116
Block 7 (Duration – w2)	1539.5	71.7%	28.502*	0.102	0.143
Block 8 (Mean F0 – w2)	1538.7	71.9%	0.734	0.103	0.144
Block 9 (Max F0 – w2)	1533.3	71.9%	5.450*	0.106	0.149
Block 10 (Min F0 – w2)	1533.3	71.9%	0.003	0.106	0.149
Block 11 (F0 Range – w2)	1533.0	72.0%	0.324	0.106	0.149
Block 12 (Intensity – w2)	1525.8	71.2%	7.177*	0.111	0.156

Note. This table presents the results of a model that compares *wh-* questions that start with two closed-class words (Wh_cc) versus statements that start with two closed-class words (S_cc) for the first (w1) and second words (w2).
*Significant at $\alpha = 0.05$.

Appendix C. Model results – Wh_cc vs S_oo

	-2LL	Data correctly Classified	χ^2 (df = 1)	R^2 Cox & Snell	R^2 Nagelkerke
Block 0 (no predictors)	913.3	62.10%			
Block 1 (Duration – w1)	828.8	70.30%	84.482*	0.116	0.157
Block 2 (Mean F0 – w1)	816.2	71.50%	12.641*	0.132	0.179
Block 3 (Max F0 – w1)	814.6	72.20%	1.604	0.134	0.182
Block 4 (Min F0 – w1)	813.5	73.00%	1.104	0.135	0.184
Block 5 (F0 Range – w1)	811.1	72.20%	2.397	0.138	0.188
Block 6 (Intensity – w1)	759.6	77.90%	51.469*	0.200	0.272
Block 7 (Duration – w2)	610.6	80.70%	148.98*	0.356	0.484
Block 8 (Mean F0 – w2)	603.8	80.80%	6.802*	0.362	0.493
Block 9 (Max F0 – w2)	602.3	81.50%	1.504	0.364	0.495

	-2LL	Data correctly Classified	χ^2 (df = 1)	R^2 Cox & Snell	R^2 Nagelkerke
Block 10 (Min F0 – w2)	598.3	81.00%	4.014*	0.367	0.500
Block 11 (F0 Range – w2)	598.2	81.30%	0.097	0.367	0.500
Block 12 (Intensity – w2)	529.1	84.00%	69.139*	0.428	0.582

Note. This table presents the results of a model that compares *wh*- questions that start with two closed-class words (Wh_cc) versus statements that start with two open-class words (S_oo) for the first (w1) and second words (w2).
*Significant at $\alpha = 0.05$.

Appendix D. Model results – Wh_cc vs S_oc

	-2LL	Data correctly classified	χ^2 (df = 1)	R^2 Cox & Snell	R^2 Nagelkerke
Block 0 (no predictors)	1691.0	68.60%			
Block 1 (Duration – w1)	1624.4	68.80%	66.555*	0.048	0.067
Block 2 (Mean F0 – w1)	1624.1	68.70%	0.346	0.048	0.068
Block 3 (Max F0 – w1)	1617.7	68.60%	6.406*	0.053	0.074
Block 4 (Min F0 – w1)	1614.1	68.90%	3.584	0.055	0.077
Block 5 (F0 Range – w1)	1611.9	68.10%	2.193	0.057	0.079
Block 6 (Intensity – w1)	1545.9	69.20%	65.961*	0.101	0.142
Block 7 (Duration – w2)	1486.0	72.20%	59.985*	0.140	0.197
Block 8 (Mean F0 – w2)	1484.2	72.00%	1.743	0.141	0.198
Block 9 (Max F0 – w2)	1465.6	73.30%	18.569*	0.153	0.215
Block 10 (Min F0 – w2)	1465.3	73.50%	0.299	0.153	0.215
Block 11 (F0 Range – w2)	1464.8	73.70%	0.552	0.153	0.215
Block 12 (Intensity – w2)	1432.7	73.70%	32.119*	0.173	0.243

Note. This table presents the results of a model that compares *wh*- questions that start with two closed-class words (Wh_cc) versus statements that start with an open-class word followed by a closed-class word (S_oc) for the first (w1) and second words (w2).
*Significant at $\alpha = 0.05$.

https://doi.org/10.1017/S0305000922000460 Published online by Cambridge University Press

Appendix E. Model results – Wh_cc vs S_co

	-2LL	Data correctly classified	χ^2 (df = 1)	R^2 Cox & Snell	Nagelkerke
Block 0 (no predictors)	1562.4	64.40%		0.026	0.035
Block 1 (Duration – w1)	1531.1	64.50%	31.293*	0.073	0.100
Block 2 (Mean F0 – w1)	1471.7	67.50%	59.389*	0.074	0.101
Block 3 (Max F0 – w1)	1470.3	67.90%	1.351	0.074	0.101
Block 4 (Min F0 – w1)	1470.3	68.10%	0.037	0.077	0.106
Block 5 (F0 Range – w1)	1466.2	67.80%	4.062*	0.099	0.136
Block 6 (Intensity – w1)	1437.0	67.40%	29.235*	0.330	0.453
Block 7 (Duration – w2)	1081.6	79.70%	355.36*	0.331	0.455
Block 8 (Mean F0 – w2)	1080.1	80.00%	1.534	0.334	0.459
Block 9 (Max F0 – w2)	1074.2	80.80%	5.916*	0.335	0.459
Block 10 (Min F0 – w2)	1073.7	80.30%	0.504	0.335	0.461
Block 11 (F0 Range – w2)	1072.2	80.00%	1.458	0.362	0.497
Block 12 (Intensity – w2)	1023.9	81.20%	48.314*	0.026	0.035

Note. This table presents the results of a model that compares *wh-* questions that start with two closed-class words (Wh_cc) versus statements that start with a closed-class word followed by an open-class word (S_co) for the first (w1) and second words (w2).
*Significant at $\alpha = 0.05$.

Cite this article: Geffen, S., Burkinshaw, K., Athanasopoulou, A., & **Curtin, S.** (2024). Utterance-Initial Prosodic Differences Between Statements and Questions in Infant-Directed Speech. *Journal of Child Language* **51,** 137–167, https://doi.org/10.1017/S0305000922000460