Say It Like You Mean It: Mothers’ Use of Prosody to Convey Word Meaning

Debora S. Herold
Indiana University-Purdue University Indianapolis, USA

Lynne C. Nygaard and Laura L. Namy
Emory University, USA

Abstract
Prosody plays a variety of roles in infants’ communicative development, aiding in attention modulation, speech segmentation, and syntax acquisition. This study investigates the extent to which parents also spontaneously modulate prosodic aspects of infant directed speech in ways that distinguish semantic aspects of language. Fourteen mothers of two-year-old children read a picture book to their children in which they labeled pictures using dimensional adjectives (e.g., big, small, hot, cold). Recordings of the mothers’ input to their children were analyzed acoustically and antonyms within each dimension were compared. Mothers modulated aspects of their prosody including amplitude and duration of target words and sentences to distinguish dimensional adjectives. Mothers appear to recruit prosody in the service of word learning.

Keywords
infant directed speech, prosody, tone of voice, word learning, word meaning

Introduction
Traditionally, research on the role of prosody in infant communication has focused on how prosodic cues like rhythm, pitch, intonation, and stress vary in Infant Directed Speech (IDS) in comparison to Adult Directed Speech (ADS). IDS consists of utterances spoken with a higher pitch, wider pitch excursions, shorter utterances, and longer pauses than typical ADS. Infants prefer to listen to this type of speech over adult directed speech from birth (Fernald, 1985, 1992; Fernald & Kuhl, 1987; Kitamura & Burnham, 2003; Pegg, Werker, & McLeod, 1992; Werker, Pegg, & McLeod, 1994).

Corresponding author:
Debora S. Herold, Indiana University-Purdue University Indianapolis, Department of Psychology, 402 N. Blackford St., LD 124, Indianapolis, IN 46202, USA
Email: dssasso@iupui.edu
In addition to simply preferring IDS, infants appear to make use of prosody in the service of learning language, suggesting that prosody serves a variety of different functions. The prosodic characteristics of IDS appear to simplify the speech stream and to aid in comprehension (see Fernald, 1984, 1991 and Snow, 1977 for reviews). Research demonstrates that when teaching new words to 12-month-olds, mothers often present words in isolation or place target words in an utterance final position and use stress in order to focus the child’s attention on the target word (Aslin, Woodward, LaMendola, & Bever, 1996; Fernald & Mazzie, 1991; Gogate, Bolzani, & Betancourt, 2006). Infants also use the stress pattern information typical of IDS to identify the boundaries of words (e.g., Echols, Crowhurst, & Childers, 1997; Johnson & Jusczyk, 2001; Thiessen, Hill, & Saffran, 2005) and to identify grammatical form class, which can help to constrain meaning (Fisher & Tokura, 1996; Shi, Werker, & Morgan, 1999; Soderstrom, Blossom, Foygel, & Morgan, 2008). Finally, for prelinguistic infants, prosody appears to be integrated with phonological information in infants’ lexical representations such that they more readily recognize familiar words when the prosody at test matches prosody during familiarization (Bortfeld & Morgan, 2010; Singh, Morgan, & White, 2004).

Fernald has argued that the role of prosody in communicative development changes over time (Fernald, 1991, 1992). In young infants, IDS serves a social function, engaging, maintaining, and modulating the infant’s attention. With age and experience, as children start to process language meaning, the distinctive prosodic aspects of IDS take on a more linguistic function, helping the child to segment words in an utterance and to comprehend word meaning. IDS appears to serve as a bootstrap to guide infants into the language learning process.

Although research on the role of IDS in communication indicates that this specialized form of speech serves functions ranging from capturing the child’s attention to segmenting words to identifying grammatical structure, the ways in which IDS may communicate information about the semantics of words have been less clear. The majority of research on the relation between prosody and meaning has focused on ways that the prosodic features of IDS may communicate information about the emotion of a speaker. By five months, infants are already responding differently to happy and sad vocal expressions (Fernald, 1991). Fernald (1992) has argued that infants detect emotional content of speech from prosodic contours via autonomic arousal responses in the infant induced by the prosodic contour. For example, if a mother shouts, “STOP!” as an infant is reaching towards a hot stove, the intensity of the tone of voice will cause the child to become startled, open her eyes, and orient herself towards the speaker, even though she may not understand the meaning of the word. Fernald has suggested that this type of response forms the basis for children’s eventual comprehension of language. Over time, the endpoint response evoked by the tone of voice used during this exclamation becomes associated with the word. The parent can exploit the child’s natural responses to different vocal tones and use this to teach the meaning of emotion-related utterances (Fernald, 1992; Lewis, 1936/1951).

Research on the prosodic characteristics of adult directed speech has focused on many of the same aspects of its contribution to communicative events. Prosodic cues communicate information about the structure of spoken language by signaling syllable, word, and phrase boundaries (Cutler, 1996; Cutler & Norris, 1988; Shattuck-Hufnagel & Turk, 1996), about the utterance type (e.g., question vs. declaration, Soderstrom, Ko, & Nezvoro, 2011), and about the emotional or attitudinal state of the speaker (Bachorowski, 1999; Scherer, 1994; Scherer, Banse, Wallbott, & Goldbeck, 1991).

Less attention has been paid to whether prosodic cues may communicate information about the content of utterances, particularly non-emotion-related utterances. A growing body of evidence has indicated that there are non-arbitrary sound-to-meaning correspondences in speech communication that may facilitate listeners’ ability to process meaning beyond valence. One type of sound-to-meaning correspondence in natural language is phonetic sound symbolism. Research indicates that
non-arbitrary phonology-to-meaning correspondences exist in spoken languages (Kita, 1997; Nuckolls, 1999), that children are sensitive to these cues to meaning (Imai, Kita, Nagumo, & Okada, 2008; Maurer, Pathman, & Mondloch, 2006), and that these sound-to-meaning correspondences can even facilitate word learning cross-linguistically (Kantartzi, Imai, & Kita, 2011; Kunihira, 1971; Nygaard, Cook, & Namy, 2009a; Yoshida, in press). These effects are not prosodic but rather reflect mappings between the phonetic features of the lexical form and its meaning. However, prosodic information also appears to serve as an independent source of information that is largely orthogonal to phonetic content but nonetheless provides additional information that facilitates the interpretation of words and sentences.

Several recent studies indicate that prosodic cues can indeed be recruited by adult and child listeners to infer the meaning of non-emotion-related utterances, even within the context of fixed phonetic content. Shintel and Nusbaum (2007; Shintel, Nusbaum, & Okrent, 2006) demonstrated both that adult speakers modify the prosodic characteristics of speech to reflect meaning, and that adult listeners readily utilize these characteristics to facilitate the processing of meaning. For example, speakers describing the trajectory (left or right) of a dot varied the speaking rate of the same sentence (e.g., “The dot is moving left.”) as a function of the rate at which the dot was moving. In a forced-choice perceptual task, listeners reliably guessed which dot display the speaker was describing (e.g., slow vs. fast) from the prosodic information alone.

Kunihira (1971) demonstrated that expressive intonation facilitated novel word interpretation by asking adult monolingual native speakers of English to choose the correct meaning for Japanese words. Participants were presented with Japanese antonym pairs (e.g., hot/cold) in one of three forms – in a printed form using English spelling conventions, spoken in a monotone voice, or spoken with expressive prosody. Although participants guessed the meaning at above chance rates in all three conditions (underscoring the availability of phonetic sound symbolism), adults in the expressive prosody condition chose the correct meaning more often than those in both the written or monotone conditions. This finding supports the idea that there are prosodic correlates (that apparently hold cross-linguistically) to meaning for both emotion and non-emotion words.

Nygaard, Herold, and Namy (2009b) conducted a similar study using nonsense words (e.g., “blicket”) recorded in IDS. Because the same word forms were recorded with different meanings (e.g., the same word was produced by speakers intending it to mean hot and again intending it to mean cold), any differences in interpretation were attributable to prosodic rather than segmental information. They reported that adult listeners could correctly infer from IDS which antonym meaning (e.g., big vs. small, hot vs. cold) to assign to a novel word on the basis of prosody alone. Nygaard et al. went on to perform acoustic analyses to investigate what features of the utterances reliably distinguished meanings. They found that some acoustic features corresponded with the overall valence of the meaning. Meanings that were considered positive in valence (based on adult ratings) such as big and yummy were characterized by higher F0, greater F0 variation, and increased amplitude while meanings with negative valence such as yucky and weak had lower F0, F0 variation, and amplitude. However, the acoustic analysis also revealed acoustic characteristics associated uniquely with particular semantic domains. For example, within the domain of size, words meaning big and tall were produced with greater amplitude and longer duration than their antonyms, small and short. These findings are the first to suggest that IDS contains reliable prosodic markers to word meaning and that adult listeners use these prosodic cues to differentiate meanings.

More recently, Herold, Nygaard, Chicos, and Namy (2011) found that children, like adults, are sensitive to these prosodic cues to meaning and are able to use prosodic information to infer word meaning. Using a modified version of the adult task described above, four- and five-year-old children listened to one of three speakers producing the same novel words using either meaningful
or meaning-neutral IDS. Both four- and five-year-olds reliably selected the correct meaning on the meaningful trials and responded randomly on the neutral trials, although four-year-olds required some additional training to succeed on the task.

This evidence suggests that prosodic cues to meaning are available in natural speech and can be readily interpreted. In the absence of any specific instructions about how to recruit prosodic cues to convey meaning, the speakers in Nygaard et al. (2009b) independently converged on the same prosodic features within a given semantic domain and systematically differentiated prosodic cues across semantic domains. This convergence suggests that such cues are common and culturally consistent across speakers in IDS. Likewise, the finding that both adult and child listeners were able to use these cues to infer meaning supports the ecological validity of these cues. However, it is unclear whether mothers spontaneously produce similar reliable prosodic cues to word meaning when using IDS in real-time interactions with their young children.

The present investigation assessed whether consistent prosodic cues, like those used by the speakers in Nygaard et al. (2009b), are employed by parents in natural speech input to their children, in the service of facilitating word interpretation. Mothers of two-year-old children were recorded producing dimensional adjectives (e.g., “Look at the big one!”) during a picture book reading interaction with their children. Parents produced each adjective in an antonym pair (e.g., “big” and “small”). Parents were also recorded reading these sentences in baseline adult-directed speech. We then performed acoustic analyses to assess differences in prosodic characteristics of mothers’ speech between meanings within and across dimensions, and also compared the relative magnitude of prosodic modulations when mothers were interacting with their children to potential modulations in the baseline adult-directed utterances. Importantly, parent participants were unaware of the specific aims of the study.

2 Method

2.1 Participants

Fourteen mother–child dyads served as participants. Mothers had a mean age of 36.53 years (range: 34.52–41.23 years). All mothers were native speakers of English and had no history of speech or hearing disorders. Their children (7 female and 7 male) had a mean age of 23.44 months (range: 21.30–25.13 months). Three additional mother–child dyads were excluded from the analysis due to failure to complete the recording session. Participants were recruited from the metro Atlanta area through direct mailings.

2.2 Materials

Six antonym pairs were selected based on their conceptual familiarity to young children. These pairs included happy/sad, hot/cold, big/small, tall/short, yummy/yucky, and strong/weak. For each antonym pair, two pairs of clip art drawings were selected that depicted their meanings (e.g., a big flower and a small flower). The 24 clip art pictures were selected from a set of candidate pictures on the basis of pilot testing to ensure that their target meanings were readily recognizable to young children (see Table 1 for sample picture stimuli). The 12 pairs of pictures (two for each of the six dimensions) were color printed, laminated, and placed in a binder to create a picture book. Each page in the book contained two pictures and sentences labeling the properties depicted in each picture. For example, under the big flower was printed the sentence, “Look at the big one.” And under the small picture was printed the sentence, “Look at the small one.” Picture order was semi-randomized for each dyad such that no two pages depicting the same meanings appeared on consecutive pages.
2.3 Procedure

Each mother and child was seated in a playroom and introduced to the experimenter who explained the study while the child and a second experimenter played on the floor with toys.

The procedure included two data collection phases. The first phase involved the collecting of baseline recordings of all sentences to be elicited in the parent–child interaction. The second phase involved a semi-naturalistic picture book reading task involving the parent and child.

2.3.1 Baseline recording. After signing the appropriate consent forms, the mother and primary experimenter moved to a separate room for baseline recording of the target sentences while the child remained in the playroom with the second experimenter. Each mother recorded 12 sentences using a Sony ECM719 lapel microphone, one for each of the 12 meanings under investigation (i.e., happy, sad, hot, cold, big, small, tall, short, yummy, yucky, strong, and weak) on a Sony digital audio tape (DAT) recorder. Each word was embedded in the carrier phrase, “Look at the (meaning) one.” The parent was given a printed list of sentences and was instructed to produce each sentence as she would normally speak to an adult. The experimenter remained in the room, directly across the table from the parent while she recorded the sentences. All sentences were presented in a random order for each parent.

After collecting the baseline ADS recordings, the experimenter showed the parent the picture book and explained that on each page she would see a pair of pictures with a sentence describing each of the pictures in the pair. The experimenter instructed the mother to read the picture book while ensuring that her child understood the meaning of the words. The parent was instructed to talk freely about the pictures as she would at home, on the condition that she should read each of the sentences written under each picture at least once at some point during the session, resulting in the production of two sample sentences for each of the 12 target dimensional adjectives depicted.

2.3.2 Picture book reading. After explaining the picture book procedure to the mother, the parent and child were reunited, and the parent was seated at a table with her child seated in a booster seat

Table 1. Sample picture stimuli.

<table>
<thead>
<tr>
<th>Happy/Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Happy/Sad Image]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hot/Cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Hot/Cold Image]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Big/Small</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Big/Small Image]</td>
</tr>
</tbody>
</table>
next to her or in her lap. The parent held the picture book and examined each page with her child. Each picture book session was digitally audio-recorded on the lapel microphone as well as videotaped. The experimenter sat unobtrusively to the side of the dyad to monitor the picture book reading. If she noticed that the parent failed to read a target sentence or that the child had been vocalizing during the mother’s reading of the sentence, she quietly asked the parent to repeat that sentence.

2.4 Acoustic analysis

Each recorded session was re-digitized on a PowerMac G3 at 44.1 kHz sampling rate and edited into separate files for acoustic analysis. The speech samples were segmented once to isolate the target sentences for each meaning and a second time to isolate only the target word (e.g., short) within each sentence. For three target meanings (sad, hot, and strong) discernable sentences in ADS were obtained for only 13 of the 14 mothers. However, each mother contributed both IDS and ADS sentences for at least 11 of the 12 meanings. For each segmented stimulus, fundamental frequency (F0), F0 variation, overall amplitude, and duration were measured using SoundScope, a speech analysis software package. These four acoustic measures were the same ones that differentiated meanings in Nygaard et al. (2009b). F0 indicates the number of cycles per second in a periodic sound and contributes to the perception of the pitch of an utterance. F0 variation was measured using the standard deviation in F0 values across the utterance. Overall amplitude reflects amount of energy produced. Duration is simply the length of the utterance which corresponds with speaking rate. Although duration is a prosodic cue that appears to reliably differentiate word meanings (Nygaard et al., 2009b), the use of duration as a metric of prosodic modulation in this study posed a special challenge given that the target words employed were not matched for length or phonemic content. To the extent that duration differences are in the direction predicted by word length, this measure does not directly implicate prosodic modulation. However, duration effects that are in the opposite direction of that predicted by word length do imply prosodic modulation. To explore the extent to which prosodic modulation was unique or especially pronounced in IDS, we compared measures that reliably differentiated meaning in IDS to those same measures for the baseline ADS sentences.

Given that we would not necessarily expect prosodic correlates to be confined to the individual words, we performed analyses both on the target word edited from the phrase and on the entire utterance, as in Nygaard et al. (2009b). Results at the sentence level largely mirrored those at the individual word level. As a result, only analyses for individual words are reported.

3 Results

To confirm that mothers were employing IDS in their interactions with their children and that their IDS differed from baseline ADS, we compared the mean values in IDS to the baseline recordings across meanings on fundamental frequency (F0), F0 variation, amplitude, and duration. We expected that, consistent with previous research on IDS, mothers would produce utterances with higher F0, higher F0 variation, higher amplitude and longer duration in IDS than baseline. Separate ANOVAs comparing Speech type (IDS vs. baseline) averaged across meaning for each acoustic property revealed main effects of Speech type for F0, $F(1, 331) = 26.86, p < .01, \eta^2_p = .08$, amplitude $F(1, 333) = 5.10, p < .05, = .02$, and duration, $F(1, 333) = 49.21, p < .01, = .13$, indicating that mothers used higher F0 ($M = 245.25$ vs. $218.99$ Hz), higher amplitude ($M = 1.19$ vs. $0.98$ rms), and longer duration ($M = 523.49$ vs. $414.13$ ms) when speaking in IDS relative to baseline. There was no difference in F0 variation between IDS and baseline. These analyses confirm that mothers employed prototypical characteristics of IDS while interacting with their children, although they did not display the greater variability in F0 characteristic of IDS in this sample.
Of particular interest for this study was whether mothers modulated their prosody to differentiate meanings within a particular dimension. To address this question, we compared the acoustic features of the two meanings within each dimension produced in IDS (e.g., tall compared to short). This measure has the benefit of being the most naturalistic way to assess how mothers vary the prosodic features of word meanings within a dimension of meaning when producing speech directed to their children. However, as we noted above, this analysis requires the comparison of utterances that include different phonemic content for the different meanings.

Figures 1 through 4 depict the within-dimension differences in IDS for each acoustic measure. Appendices A and B report the raw values for each measure in IDS and baseline respectively. Repeated measures t-tests within each antonym pair revealed no differences between meanings on measures of either F0 or F0 variation with the exception of a lower F0 for strong than weak, \( t(13) = 2.32, p < .05, d = 0.69 \). Because F0 and F0 variation remained largely constant across meanings within IDS, amplitude and duration were the main features that differentiated meanings within each dimension.

In IDS, happy, \( t(13) = 2.17, p < .05, d = 0.54 \), tall, \( t(13) = 2.84, p < .05, d = 0.64 \), and strong, \( t(13) = 3.33, p < .01, d = 1.00 \), were all produced with higher amplitude than were their respective antonym pairs. Amplitude measures failed to differentiate meanings for the hot/cold, big/small, and yummy/yucky dimensions. Individual patterns analysis revealed consistent prosodic modulation across mothers. Twelve of 14 mothers employed a higher amplitude for happy than sad. Similarly 12 of 14 mothers produced tall with a higher amplitude than short, and 13 of 14 mothers produced strong with a higher amplitude than weak.

Analysis of the duration measure revealed that in IDS, happy, \( t(13) = 4.48, p < .01, d = 0.89 \), hot, \( t(13) = 5.69, p < .01, d = 1.58 \), big, \( t(13) = 5.35, p < .01, d = 1.57 \), short, \( t(13) = 2.67, p < .05, d = 0.82 \), yucky, \( t(13) = 2.70, p < .05, d = 0.33 \), and weak, \( t(13) = 8.96, p < .01, d = 2.35 \), were all produced with shorter duration than were their antonyms. The difference patterns observed were consistent across mothers for each meaning. Twelve of 13 mothers employed a shorter duration to
distinguish happy from sad, all 14 mothers employed a shorter duration to distinguish hot from cold, 13 of 14 used a shorter duration for big than small, 11 of 14 produced short with a shorter duration than tall, 11 of 14 produced yucky with a shorter duration than yummy, and all 14 mothers produced weak with a shorter duration than strong.

**Figure 2.** Mean F0 variation in IDS for each meaning within each antonym pair. Error bars represent ±1 standard error.

**Figure 3.** Mean amplitude in IDS for each meaning within each antonym pair. Error bars represent ±1 standard error. * denotes significant difference between 1st and 2nd meaning at $p < .05$. 
Interestingly, in two cases, happy/sad and tall/short, the duration measure varied between meanings in a direction opposite to that predicted by word length. These effects, evident at both the word and sentence level, provide evidence that the differences were not exclusively due to variation across items in phonemic content. In the other cases (e.g., hot shorter than cold), we cannot definitively distinguish the effects of phonemic content from prosodic modulation because duration differences are consistent with word length. The combination of effects for amplitude and duration provided some evidence for unique acoustic profiles employed to modulate individual dimensions of meaning. For example, mothers produced happy with higher amplitude and shorter duration than sad. However, tall and strong were each produced with higher amplitude and longer duration than their antonyms. The remaining three dimensions (hot/cold, big/small, and yummy/yucky) were differentiated only on duration measures which might, of course, have been accounted for by inherent differences in word duration resulting from different phonemic content.

To address whether the differences in duration observed between antonyms are unique or exaggerated in IDS relative to ADS, we evaluated whether the duration differences between antonyms observed in IDS were comparable to those observed in the baseline recordings. For duration measures we conducted a 2 (speech register: IDS vs. ADS baseline) × 2 (Antonym meaning) ANOVA for each antonym pair. These analyses revealed that although many of the duration differences between antonyms were also apparent in the baseline measures, as evidenced by the main effects of meaning for five of the six dimensions, the interactions observed reveal that the duration effect sizes in IDS differed reliably from those in the baseline utterances for four of the six dimensions. Main effects of speech register, which indicate significant differences in duration for IDS relative to ADS, were observed for all six dimensions: happy/sad, $F(1, 52) = 11.99, p < .01, = .19$, hot/cold, $F(1, 52) = 20.72, p < .01, = .29$, big/small, $F(1, 52) = 5.37, p < .05, = .09$, tall/short, $F(1, 52) = 8.54, p < .01, = .14$, yummy/yucky, $F(1, 52) = 15.4, p < .01, = .23$, and strong/weak, $F(1, 51) = 18.74, p < .01, = .27$. Main effects of meaning, which indicate differences in duration between meaning across speech registers, were
observed for the happy/sad, $F(1, 52) = 7.35, p < .01, = .12$, hot/cold, $F(1, 52) = 21.57, p < .01, = .29$, big/small, $F(1, 52) = 55.38, p < .01, = .52$, and strong/weak dimensions, $F(1, 51) = 56.87, p < .01, = .53$, with a marginal effect observed for the tall/short dimension, $F(1, 52) = 2.62, p = .12, = .05$. Interactions that support a larger magnitude of difference between IDS and ADS were observed for hot/cold, $F(1, 52) = 8.74, p < .01, = .14$, tall/short, $F(1, 52) = 4.25, p < .05, = .08$, and strong/weak dimensions, $F(1, 51) = 7.39, p < .01, = .13$, with a marginal effect observed for the yummy/yucky dimension, $F(1, 52) = 2.49, p = .12, = .05$. These findings indicate that parents were modulating the duration of the individual words in IDS relative to their baseline speaking rate for these target words, although in several cases reliable duration differences were observed in ADS as well.

4 Discussion

This semi-naturalistic study provides the first demonstration that mothers systematically vary the prosodic contours of their speech contrastively to differentiate word meanings. Despite the fact that they were unaware of the aims of our study, parents spontaneously recruited prosodic information in their interactions with their children to differentiate word meaning. Previous research has clearly demonstrated prosodic correlates to valence (i.e., positive versus negative emotion), and the present results confirm that parents produce acoustic cues to valence in IDS (Scherer et al., 1991). For example, as found in previous research, mothers in the present study used a higher amplitude in IDS when describing the “happy” pictures than when describing the “sad” pictures. Utterances describing happy pictures were also characterized by a shorter duration than were those describing sad pictures. This acoustic profile was unique to these words, suggesting that the prosodic variation employed for the other words was not a function of valence. The finding that mothers employ valence cues when referring to happiness and sadness is a novel contribution in and of itself, demonstrating that mothers employ prosodic cues to emotion not only when experiencing the emotions themselves (Fernald, 1992), but also when referring to valenced meanings.

More importantly, we also found that mothers also differentiated non-emotional antonyms via prosody, captured primarily by amplitude and duration. These patterns were distinct from those used to differentiate happy from sad. These acoustic profiles varied across dimensions of meaning but were highly consistent across speakers within each dimension. This paradigm was not, of course, entirely naturalistic in that we elicited these particular target sentences as part of the ongoing interaction between mothers and their children. Nonetheless, mothers spontaneously employed prosodic modulation to contrast which word mapped to which picture within each antonym pair. It is possible that the structure of the task (contrasting dimensional adjectives) implicitly encouraged mothers to employ or exaggerate prosodic modulation of the antonyms within each pair. However, mothers were highly consistent in recruiting particular prosodic modulations within the context of particular antonym pairs and these prosodic modulations varied across pairs, suggesting that this is a natural and consistent feature of infant-directed interactions.

The findings that some of the effects were also observed in the baseline recordings lend support to the idea that this may not be unique to IDS but rather that speakers may recruit prosody to convey meaning across speech registers.

There are several possible explanations for how and why mothers employed prosody in the service of facilitating children’s word-to-referent mapping. One possibility is that these prosodic correlates to meaning occur automatically in native English speakers as a result of having learned language-specific prosodic conventions but are not necessarily strategically recruited by parents in the service of facilitating word learning. That is, perhaps English language conventions include particular prosodic modulations of individual lexical items that are standardized and employed broadly by speakers in general. The consistency in prosodic patterns observed across mothers in both IDS and ADS baseline
sentences is consistent with this interpretation, although the typically more exaggerated prosodic modulation in IDS than in baseline is not, suggesting that this cannot be the entire explanation for our findings.

A second possibility is that prosodic correlates to meaning are a by-product of simulation or embodiment of the meanings (Niedenthal, Barsalou, Rie, & Krauth-Gruber, 2005). For example, perhaps the greater amplitude employed in the case of *strong* than *weak* may result from simulating the experience of being a stronger, heartier person whose voice would also resonate more loudly. The consistency displayed across speakers is consistent with this account as well. An embodiment account might also predict the more pronounced prosodic modulation in IDS than at baseline – not because of the change in register from ADS to IDS but because the picture book reading context may have elicited a richer simulation of the concepts depicted than the baseline sentence reading context. That many of the effects observed in IDS were also found in the baseline ADS recordings, despite the relatively impoverished eliciting conditions, suggests that this prosodic modulation is not a uniquely IDS-based phenomenon.

However, the fact that the effects observed were generally more robust in IDS than in baseline also suggests a third possibility, that mothers are intentionally and strategically exaggerating prosodic cues to facilitate their children’s comprehension. Future work in which the eliciting contexts for IDS and ADS are well matched in semantic richness and social interaction will be needed to disambiguate these latter two accounts.

Our amplitude modulation measure is a straightforward index of prosodic modulation, and reliably distinguished meanings within three of the six antonym pairs. Our duration measure was more difficult to interpret. The variability in word length across words within an antonym pair renders it difficult to isolate the unique effects of prosody with respect to the duration measure. Only in two cases, *happy/sad* and *tall/short*, in which the duration effects were in the opposite direction of those predicted by word length, can we definitively implicate prosodic variation as a contributor. It is interesting, for example, that *tall/short* displayed this effect but not *big/small*, although this is consistent with speakers’ use of larger differences in duration when distinguishing *tall* from *short* than when distinguishing *big* from *small* in previous findings by Nygaard et al. (2009b). Nonetheless, the fact that at least some of the items displayed longer durations for shorter words suggests that duration is an aspect of prosody that is employed by mothers to signal reference. This is consistent with findings by Nygaard et al., (2009b) that duration is a reliable index of word meaning for at least some antonym contrasts.

It is unclear whether the use of prosodic modulation to convey semantic information observed in this study is a generalizable phenomenon that is consistent and comprehensive across word meanings. The contrastive use of antonyms in this study may have been particularly likely to elicit such prosodic modulation. Extension of this paradigm to investigate non-contrastive use of the same adjectives, to assess whether the same prosodic modulations are used when labeling objects or events that possess these characteristics (e.g., when labeling a big vs. a small dog as a “dog”), and to identify prosodic correlates of other aspects of meaning will all provide a clearer picture of the pervasiveness of the use of prosodic cues to semantics.

These findings are the first to demonstrate that mothers employ consistent and reliable prosodic cues to differentiate word-to-referent mappings when speaking to their young children. This important aspect of speech not only contributes to the segmentation of speech or to acquiring information about the emotion of a speaker but also conveys and differentiates word meaning. This raises the possibility that children may be sensitive to these prosodic cues and able to make use of them in the service of learning novel words. Herold et al. (2011) have shown that preschool children can reliably map novel words to their intended referents on the basis of prosody alone. An important question for future research is whether children, at the earliest stages of word learning, can capitalize on available prosodic cues to facilitate novel word interpretation.
The current study adds to a growing body of research suggesting that prosodic information serves multiple important functions early in development. Prosody acts to simplify the speech stream, modulate infant attention, signal emotion, and aid in syntax acquisition. The present findings imply that parental prosody serves an additional function, conveying reliable information about word meaning.

Acknowledgements

We thank the members of the Language and Learning Laboratory at Emory University for help with data collection. Many thanks to all of the parents and children who participated in this research.

Note

1. F0 was derived using a peak-picking algorithm. F0 variation was calculated from the standard deviation of frame-by-frame F0 values. Utterances that were longer in duration contributed more values to the statistic than did those that were shorter in duration.

References


## Appendix A

Acoustic measurements of English words in sentences produced with IDS

<table>
<thead>
<tr>
<th></th>
<th>F0 (Hz)</th>
<th>F0 SD (Hz)</th>
<th>Amplitude (rms)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Happy/Sad dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>255.75</td>
<td>71.34</td>
<td>1.21</td>
<td>486.27</td>
</tr>
<tr>
<td>Sad</td>
<td>235.56</td>
<td>63.08</td>
<td>0.75</td>
<td>590.27</td>
</tr>
<tr>
<td><strong>Hot/Cold dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot</td>
<td>257.43</td>
<td>60.78</td>
<td>1.41</td>
<td>418.30</td>
</tr>
<tr>
<td>Cold</td>
<td>231.44</td>
<td>52.35</td>
<td>1.67</td>
<td>710.54</td>
</tr>
<tr>
<td><strong>Big/Small dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>272.71</td>
<td>58.15</td>
<td>1.15</td>
<td>343.20</td>
</tr>
<tr>
<td>Small</td>
<td>265.64</td>
<td>51.71</td>
<td>0.97</td>
<td>557.10</td>
</tr>
<tr>
<td><strong>Tall/Short dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tall</td>
<td>249.72</td>
<td>61.64</td>
<td>1.87</td>
<td>558.81</td>
</tr>
<tr>
<td>Short</td>
<td>267.89</td>
<td>59.35</td>
<td>1.11</td>
<td>456.69</td>
</tr>
<tr>
<td><strong>Yummy/Yucky dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yummy</td>
<td>219.29</td>
<td>67.71</td>
<td>1.11</td>
<td>517.71</td>
</tr>
<tr>
<td>Yucky</td>
<td>228.28</td>
<td>55.57</td>
<td>1.12</td>
<td>477.28</td>
</tr>
<tr>
<td><strong>Strong/Weak dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>214.54</td>
<td>75.80</td>
<td>1.39</td>
<td>744.83</td>
</tr>
<tr>
<td>Weak</td>
<td>244.75</td>
<td>67.23</td>
<td>0.52</td>
<td>420.63</td>
</tr>
</tbody>
</table>

## Appendix B

Acoustic measurements of English words in sentences produced with ADS

<table>
<thead>
<tr>
<th></th>
<th>F0 (Hz)</th>
<th>F0 SD (Hz)</th>
<th>Amplitude (rms)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Happy/Sad dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>224.49</td>
<td>75.14</td>
<td>0.96</td>
<td>425.92</td>
</tr>
<tr>
<td>Sad</td>
<td>207.60</td>
<td>63.60</td>
<td>1.18</td>
<td>466.24</td>
</tr>
<tr>
<td><strong>Hot/Cold dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot</td>
<td>227.55</td>
<td>46.97</td>
<td>1.41</td>
<td>356.98</td>
</tr>
<tr>
<td>Cold</td>
<td>214.11</td>
<td>51.05</td>
<td>0.98</td>
<td>421.88</td>
</tr>
<tr>
<td><strong>Big/Small dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>226.11</td>
<td>63.08</td>
<td>0.83</td>
<td>268.38</td>
</tr>
<tr>
<td>Small</td>
<td>210.39</td>
<td>54.55</td>
<td>0.98</td>
<td>494.80</td>
</tr>
<tr>
<td><strong>Tall/Short dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tall</td>
<td>198.14</td>
<td>45.37</td>
<td>1.34</td>
<td>420.89</td>
</tr>
<tr>
<td>Short</td>
<td>223.74</td>
<td>44.12</td>
<td>0.85</td>
<td>433.12</td>
</tr>
<tr>
<td><strong>Yummy/Yucky dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yummy</td>
<td>219.21</td>
<td>66.05</td>
<td>0.86</td>
<td>380.46</td>
</tr>
<tr>
<td>Yucky</td>
<td>214.23</td>
<td>73.16</td>
<td>0.87</td>
<td>418.71</td>
</tr>
<tr>
<td><strong>Strong/Weak dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>221.91</td>
<td>64.35</td>
<td>0.90</td>
<td>522.15</td>
</tr>
<tr>
<td>Weak</td>
<td>240.39</td>
<td>84.75</td>
<td>0.51</td>
<td>369.73</td>
</tr>
</tbody>
</table>