Acoustic differences between humorous and sincere communicative intentions

Elena Hoicka1* and Merideth Gattis2

1Psychology, School of Natural Sciences, University of Stirling, UK
2School of Psychology, Cardiff University, UK

Previous studies indicate that the acoustic features of speech discriminate between positive and negative communicative intentions, such as approval and prohibition. Two studies investigated whether acoustic features of speech can discriminate between two positive communicative intentions: humour and sweet-sincerity, where sweet-sincerity involved being sincere in a positive, warm-hearted way. In Study 1, 22 mothers read a book containing humorous, sweet-sincere, and neutral-sincere images to their 19- to 24-month-olds. In Study 2, 41 mothers read a book containing humorous or sweet-sincere sentences and images to their 18- to 24-month-olds. Mothers used a higher mean F0 to communicate visual humour as compared to visual sincerity. Mothers used greater F0 mean, range, and standard deviation; greater intensity mean, range, and standard deviation; and a slower speech rate to communicate verbal humour as compared to verbal sweet-sincerity. Mothers used a rising linear contour to communicate verbal humour, but used no specific contour to express verbal sweet-sincerity. We conclude that speakers provide acoustic cues enabling listeners to distinguish between positive communicative intentions.

Despite increasing evidence that children are able to distinguish the communicative intentions of others as early as 1 year, we still have little idea of how they do so. The early and relatively rapid emergence of these abilities has motivated nativist accounts. Leslie, for instance, proposed that early understanding of others’ mental states, such as pretence and false belief, is innate and modular in nature, stating ‘It is hard to see how perceptual evidence could ever force an adult, let alone a young child, to invent the idea of unobservable mental states’ (1987, p. 422).

Alternate proposals emphasize the richness of environmental cues to help children differentiate communicative intentions. Such proposals are consistent with the claim that children are biologically prepared to detect communicative intentions in general, but focus on specific environmental cues that help children distinguish amongst various possible communicative intentions. For example, 14- to 18-month-old children distinguish intentions from accidents, copying intentional actions more than accidental.

*Correspondence should be addressed to Elena Hoicka, Psychology, School of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK (e-mail: elena.hoicka@stir.ac.uk).

DOI:10.1111/j.2044-835X.2011.02062.x
actions (Carpenter, Akhtar, & Tomasello, 1998). Recent studies of how children do so have demonstrated an influence of vocal intonation on children’s inferences about the intentional status of actions (Sakkalou & Gattis, 2011).

This paper reports two studies investigating environmental cues to communicative intentions. Both studies focus on the acoustic cues that mothers provide to children when communicating two positive communicative intentions: humorous intentions and warm-hearted, sincere intentions (hereafter called sweet-sincere). Distinguishing humorous and sweet-sincere intentions involves additional complexity beyond simply differentiating positive and negative intentions such as approval and prohibition or intentions and accidents. Differentiating humour and sweet-sincerity cannot rely on a contrast between positive and negative, because both are positive. Humour and sweet-sincerity do differ in other significant ways, however, since communicating humorous intentions also requires the speaker to convey that she did the wrong thing intentionally, while communicating sweet-sincerity requires the speaker to convey that she did the right thing intentionally (Leekam, 1991; Hoicka & Gattis, 2008). Additionally, while other communicative intentions involve a lack of sincerity or literal meaning (e.g., pretence, lying, and metaphor), humour is present significantly more often in books aimed at 1- to 2-year-olds than these other types of insincere or non-literal speech (Hoicka, Jutsum, & Gattis, 2008).

Previous research demonstrates that when reading books to children, parents provide lexical cues to differentiate between humour and sincerity. Hoicka et al. (2008) found that parents use more abstract language when reading humorous book content to children. In addition, parents reading humorous book content also used more disbelief statements, clearly indicating to children that book content was not an accurate depiction of the social or physical world, and thereby encouraging children to disbelieve what was read. Hoicka et al. (2008) suggested that parents use disbelief statements to cue toddlers that the humorous utterance is wrong, or in other words incongruous, and use abstract language to encourage the child to think in an abstract way in order to resolve the incongruity.

An important limitation of lexical cues to communicative intentions is that multiple communicative intentions can be present in a single utterance. For example, a storybook might contain a verbal joke such as ‘The ducks say moo’, and when reading aloud a parent needs to convey both the semantic meaning and the humorous intention at the same time. When this happens, it is not possible for parents to provide lexical cues to the humorous intention at the same time as reading storybook text, and parents may need to provide other cues. Additionally, toddlers’ abilities to utilize lexical cues to humour are limited by the size of their receptive vocabularies. Parents may thus need to show a change in footing, suggesting that they are switching to a non-serious frame (e.g., Goffman, 1981). In particular, parents may need to express contextualization cues that implicitly express a non-serious frame, and at the same time are understandable for toddlers (e.g., Gumperz, 1982). For these reasons, extra-lexical acoustic cues may be an important component of early communication, helping younger children overcome the limitations of lexical cues, which are suggested to be involved in signalling a change in footing, and serve as contextualization cues (Goffman, 1981; Gumperz, 1982).

Previous studies have shown that mothers do use acoustic cues to signal communicative intentions to toddlers and infants. For example, Fernald (1989) presented adults with content-filtered speech samples taken from mothers expressing messages such as
Humorous and sincere communicative intentions

approval and prohibition. People were more accurate at judging the communicative intention when the speech samples were taken from infant-directed speech (IDS) compared to adult-directed speech. Fernald concluded that maternal speech to infants emphasizes melody, or intonation contour, in order to make communicative intentions more accessible to infants. Other research has shown that maternal speech uses contour to highlight communicative intentions such as attention bids, approval, or comfort (Katz, Cohn, & Moore, 1996). Other acoustic features of speech, such as fundamental frequency (F0, pitch), distinguish communicative intentions such as pretend play from more commonplace intentions such as feeding (Reissland & Snow, 1996). Furthermore, adults use acoustic cues to signal a form related to humorous intentions: ironic intentions. Adults expressing irony compared to neutral speech use a higher mean F0, larger F0 standard deviation, wider F0 range, higher mean energy (loudness), wider energy range, and faster speech rate (Anolli, Ciceri, & Infantino, 2000). Given that past research has demonstrated that mothers cue infants and toddlers to communicative intentions with acoustic features of speech, and that adults use acoustic features to cue humour-related intentions, we wanted to know whether parents would use acoustic cues to differentiate two positive communicative intentions: humour and sweet-sincerity. We included a neutral-sincere baseline in Study 1 because previous studies have demonstrated that a neutral condition provides a useful general baseline for comparisons (e.g., Anolli et al., 2000; Reissland & Snow, 1996). Our primary focus, however, was the more informative comparison of an emotionally positive sweet-sincere condition, which controls for the positive emotion inherent to humour, since positive emotions are themselves conveyed with acoustic cues (e.g., Scherer, 1986; Banse & Scherer, 1996).

The two studies reported here investigate whether mothers produce significant vocal acoustic cues that could allow a toddler to distinguish between humorous and sincere communicative intentions. For a complete comparison, Study 1 investigated the contrast between humorous intentions and two types of sincerity: sweet-sincere and neutral-sincere. Study 2 investigated the contrast between humorous intentions and sweet-sincere intentions. Based on the above review, seven summary features were considered because these features have been found to characterize distinct vocal profiles for different communicative intentions: mean F0, F0 range, F0 standard deviation, mean intensity (loudness), intensity range, intensity standard deviation (Anolli et al., 2000, measured energy rather than intensity, but these represent a conversion of the same measurement), speech rate, and intonation contour (Katz et al., 1996).

In the first study, mothers read a book to their 19- to 24-month-olds containing neutral text describing a baby’s day, accompanied by images that were either humorous or sincere (sweet-sincere or neutral-sincere). Humorous images were incongruous and invoked positive feelings. Sincere images were congruous and invoked either sweet, positive feelings (sweet-sincere), or neutral feelings (neutral-sincere). In the second study, we created two books to allow a more controlled comparison, and focused on humorous and sweet-sincere intentions only. Mothers read a book either containing humorous sentences (verbal humour), or sweet-sincere sentences (verbal sincerity) accompanied by images. The designs of the two studies allowed us to compare acoustic features of speech conveying different communicative intentions. Our analyses focused on whether parents use vocal acoustic cues to differentiate amongst positive communicative intentions.
STUDY 1
In the first study, each mother read a book to her child that contained neutral text describing a baby’s day alongside humorous and sincere images. This study explored the vocal acoustic cues that mothers use to express humorous versus sincere (both sweet-sincere and neutral-sincere) intentions. This study thus explored whether mothers cue their toddlers to visual (not verbal) humour. This is comparable to the study by Reissland and Snow (1996) in which they examined vocal acoustic cues that might cue a child to a pretend situation, but allows greater control because the acoustic cues are produced when text is read aloud. Furthermore, a within-subjects design creates a natural context similar to daily life. In everyday life, mothers switch between communicative intentions while interacting with their children.

Method
Participants
Twenty-seven mother–toddler dyads participated. Five mother–toddler dyads were dropped from the study because the children were restless. Toddlers ranged in age from 19 to 24 months with a mean age of 22 months and 1 day and a standard deviation of 1 month and 23 days. Eleven of the toddlers were girls and 11 were boys. Mother–toddler dyads were recruited from postcards sent out to playgroups and nurseries.

Materials
The book used in this study was One Gorgeous Baby by Oborne and Godon (2004). This was a counting book with 12 pages total (see Appendix A) containing sentences that were not very humorous, sweet-sincere, or neutral-sincere on their own, but in combination with the pictures, could be construed as humorous, sweet-sincere, or neutral-sincere. One humorous combination was the text ‘Four clean nappies’ with a picture of a child throwing nappies/diapers. One sweet-sincere combination was the text ‘Five cuddly toys’ with a picture of a child holding and playing with toys. One neutral-sincere combination was the text ‘Six shiny buttons’ with a picture of a child in a coat with six buttons. Book pages are shown in Appendix A. A digital video camera, a DV camcorder, and a mixing board were used to record the reading sessions. The testing room contained a sofa that had a large seat for a parent and a small, elevated seat for a toddler so that the toddler sat approximately at eye level with his or her parent. An omni-directional Beyer Boundary microphone, model MPC66VC, with a frequency response of 50 Hz to 20 kHz, was placed on the wall at mouth level behind the sofa. As the microphone needed to feed into the video recording, we chose a boundary microphone as this would provide better sound quality than a wireless microphone. Software included speech-processing program PRAAT (Boersma & Weenink, 2008).

To identify the communicative content of book pages, we first categorized pages as humorous, sweet-sincere, or neutral-sincere (see Appendix A for book pages). We then asked a separate group of 20 parents (some of whom participated in Study 1) to rate the humour and sweetness of book pages on a nine-point scale. A three-way (communicative intention: humorous, sweet-sincere, and neutral-sincere) General Linear Model (GLM) for parental ratings of page humour violated Mauchly’s test of Sphericity, $W(2) = .621, p < .025$. After a Greenhouse–Geisser adjustment, an overall difference was found, $F(1.45, 27.55) = 14.53, p < .001$. A planned Helmert contrast found that the humorous pages
Humorous and sincere communicative intentions

$M = 4.86, SD = 2.88$ were significantly more humorous than other page types (sweet-sincere and neutral-sincere), $F(1, 19) = 18.21, p < .001$, while there was no difference in humour ratings for sweet-sincere ($M = 2.63, SD = 2.45$) and neutral-sincere ($M = 3.08, SD = 2.55$) pages. A three-way (communicative intention) GLM for parental rating of page sweetness was significant, $F(2, 18) = 30.81, p < .001$. A planned Helmert contrast found that the sweet-sincere pages ($M = 6.86, SD = 1.87$) were significantly more sweet than other page types (humorous and neutral-sincere), $F(1, 19) = 64.34, p < .001$, while there was no difference in sweetness ratings for humorous ($M = 4.25, SD = 2.23$) and neutral-sincere ($M = 5.05, SD = 1.92$) pages.

**Design**

The independent variable for Study 1 was the communicative intention of the utterance: humorous or sincere. Sincere communicative intentions were further subdivided into sweet-sincere and neutral-sincere. This was done so that we had two types of controls. Because humour is emotionally positive in and of itself, the sweet-sincere control was an emotionally positive control to ensure that any humorous cues found were not due to positive emotional prosody alone, but to humour more specifically. The neutral-sincere control acted as a general neutral baseline. The dependent variables included the following summary features: mean fundamental frequency ($F_0$), standard deviation of $F_0$, and $F_0$ range, in semitones (st); mean intensity, standard deviation of intensity, and intensity range, in decibels (dB); and speech rate (syllables per second). The intonation-dependent variable was the shape of the contour (linear, quadratic, or neither).

**Procedure**

Before the study, mothers were asked to read a book to their child as they would at home. The mother was given the book to read alone before reading aloud to the child. Mother-toddler dyads were brought into the lab and asked to sit on the sofa and to read the book when ready.

**Coding**

For each mother, the six utterances (two humorous, two sweet-sincere, and two neutral-sincere) were re-digitized from DV video recording to wav files using PRAAT speech-processing software at a sampling rate of 44 kHz, and 16-bit quantization. Each utterance was visually and auditorily inspected and edited in PRAAT so that each wav file contained only the utterance and no excess segments either before or after the focal utterance (including silent segments). Using an autocorrelation method (Boersma, 1993), $F_0$ measures were coded within a range of 100 to 600 Hz, using a 30-ms Hanning window, and a 7.5-ms time step. The $F_0$ range was based on Gussenhoven’s (2004) report that women’s voices are generally in the 100–600 Hz range. As well as it being unlikely that mothers would make utterances outside this range, measuring pitch at a higher level might capture pitch-doubling errors by the software rather than actual pitch height. Mean, range, and standard deviation of $F_0$ of each utterance were then calculated from the voiced segments, and measured in semitones (st). The mean $F_0$ was calculated by averaging the $F_0$ values at each time step in an utterance in semitones. The $F_0$ range was calculated by subtracting the lowest $F_0$ measured across all time steps from the highest $F_0$ measured across all time steps within an utterance, in semitones. The $F_0$
standard deviation was measured by calculating the standard deviation of the $F_0$ values across all time steps of an utterance in semitones. The $F_0$ range differed from the $F_0$ standard deviation in that the $F_0$ range gives an overall pitch difference across the entire utterance, that is, the difference between the highest and lowest pitch, while the $F_0$ standard deviation reflects how much the pitch varies (goes up and down) throughout the utterance.

Mean, standard deviation, and range of intensity were measured in decibels (dB) with a Gaussian analysis window for voiced speech with a pitch floor of 100 Hz, using an 8-ms time step, with mean pressure subtracted. The mean intensity was calculated by returning the mean of the intensity curve in decibels with the following equation

$$\frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} x(t) dt$$

where $t_1$ is the start time, $t_2$ is the end time, and $x(t)$ is the intensity in decibels at a time point (Boersma & Weenink, 2008). The intensity range was calculated by subtracting the lowest intensity measured across all time steps from the highest intensity measured across all time steps within an utterance, in decibels. The intensity standard deviation was calculated using the equation,

$$\sqrt{\frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} dt(x(t) - \mu)^2}$$

where $t_1$ is the start time, $t_2$ is the end time, $x(t)$ is the intensity in decibels at a time point, and $\mu$ is the mean (Boersma & Weenink, 2008).

As each wav file was already edited to contain only the utterance, and no excess segments, the total utterance time was coded in PRAAT by querying the total time for each corresponding wav file. Speech rate was coded by dividing the number of syllables in the utterance by the total utterance time in seconds and milliseconds.

Scores for both utterances for each communicative intention (humorous, sweet-sincere, and neutral-sincere) were averaged for each mother on all $F_0$, intensity, and speech rate measures. Where only one phrase expressing a specific communicative intention was said by a mother, that phrase was used on its own in the final analyses.

To code the intonation contours, the coder visually and auditorally inspected the pitch contour in order to identify each syllable. For each stressed syllable (those in bold, e.g., four clean nappies, see Appendix A), the mean $F_0$ was calculated by averaging the $F_0$ values across all time steps in that syllable in semitones. Twenty-four percent of the utterances were coded by a second coder, and the correlation between mean $F_0$ for each syllable was $r = 0.94$, $n = 79$, $p < .01$. We then ran Mixed Models to test whether humorous utterances had a specific shape (e.g., linear, i.e., fall or rise, or quadratic, a.k.a. bell shape, i.e., fall rise, or rise fall), and whether the humorous contours were different from sweet-sincere and neutral-sincere contours.

Results

Mothers’ head position in relation to the microphone was coded for each utterance in case this affected the results of the acoustic analyses. Three percent of mothers’ heads were turned at 180° from the microphone; 40% were turned at 135°; and 57% were
Humorous and sincere communicative intentions

Table 1. Means and standard deviations for summary features between humorous, and sincere (sweet-sincere, neutral-sincere) communicative intentions in Study 1

<table>
<thead>
<tr>
<th>Feature</th>
<th>Humorous</th>
<th>Sweet-sincere</th>
<th>Neutral-sincere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F0 (st)**</td>
<td>17.74</td>
<td>15.22</td>
<td>15.32</td>
</tr>
<tr>
<td>SD</td>
<td>(3.13)</td>
<td>(1.55)</td>
<td>(2.56)</td>
</tr>
<tr>
<td>F0 standard deviation (st)</td>
<td>4.14</td>
<td>4.03</td>
<td>4.08</td>
</tr>
<tr>
<td>SD</td>
<td>(1.49)</td>
<td>(0.90)</td>
<td>(1.45)</td>
</tr>
<tr>
<td>F0 range (st)</td>
<td>16.32</td>
<td>16.27</td>
<td>15.43</td>
</tr>
<tr>
<td>SD</td>
<td>(5.37)</td>
<td>(3.69)</td>
<td>(4.81)</td>
</tr>
<tr>
<td>Mean intensity (dB)</td>
<td>44.67</td>
<td>44.17</td>
<td>44.54</td>
</tr>
<tr>
<td>SD</td>
<td>(2.26)</td>
<td>(1.69)</td>
<td>(2.01)</td>
</tr>
<tr>
<td>Intensity standard deviation (dB)</td>
<td>0.84</td>
<td>0.62</td>
<td>0.84</td>
</tr>
<tr>
<td>SD</td>
<td>(1.08)</td>
<td>(0.99)</td>
<td>(1.15)</td>
</tr>
<tr>
<td>Intensity range (dB)</td>
<td>3.57</td>
<td>2.82</td>
<td>3.74</td>
</tr>
<tr>
<td>SD</td>
<td>(4.04)</td>
<td>(3.84)</td>
<td>(4.63)</td>
</tr>
<tr>
<td>Speech rate (syllables/second)</td>
<td>2.32</td>
<td>3.91</td>
<td>2.42</td>
</tr>
<tr>
<td>SD</td>
<td>(0.67)</td>
<td>(0.41)</td>
<td>(0.70)</td>
</tr>
</tbody>
</table>

Note. **p < .01

turned at 90°. Mothers did not noticeably move their heads between utterances. A three-way (communicative intention: humorous, sweet-sincere, and neutral-sincere) Mixed Model found no difference between head positions in the humorous, sweet-sincere, and neutral-sincere pages.

Preliminary analyses found no effect of child age (under or over the mean age of 22 months), or child’s gender, so these variables were not included in the final analyses. Mixed models were used to allow all mothers’ available data to be used. For two mothers, data were missing for both neutral phrases due to noise (emanating from the child) or the mother not having said the phrase. Due to experimental error, the last seven participants’ intensity was recorded at a different baseline than the first 15, so date of testing was also used as a variable in intensity Mixed Models to factor out additional model error. Analyses found effects of date of testing (later-tested participants were louder than earlier-tested participants) but did not otherwise affect the analyses. Noisy data were still used for the speech rate and intonation contour analyses (although noisy syllables were left uncoded for contour analyses).

Summary features

Table 1 shows the means and standard deviations for mean F0, F0 standard deviation, F0 range, mean intensity, intensity standard deviation, intensity range, and speech rate. To evaluate differences between the three communicative intentions, three-way (communicative intention) within-subjects Mixed Models were run for F0 mean, standard deviation and range, as well as speech rate. An effect of communicative intention on mean F0 was found, F(2, 58) = 5.35, p < .01, Pseudo-R² = .171. Planned Helmert contrasts found that humorous utterances were said with a significantly higher mean F0, F(1, 58) = 9.83, p < .01 than sincere utterances (sweet-sincere and neutral-sincere). No differences were found between sweet-sincere and neutral-sincere utterances for mean
F0. No other effects of communicative intention were found. Three (communicative intention) × 2 (date of recording) within-between subjects Mixed Models were run for intensity mean, standard deviation, and range. No effects of communicative intention were found.

**Intonation contours**

A 3(communicative intention) × 3(syllable: mean F0 of stressed syllable 1, 2, 3) within-subjects Mixed Model found an effect of communicative intention, $F(2, 174) = 6.64, p < .01$, and an effect of syllable, $F(2, 174) = 6.54, p < .01$. The variance explained by the entire model was $\text{Pseudo-}R^2 = .145$. Planned contrasts found that humorous utterances were significantly higher in mean F0 overall than sincere utterances (sweet-sincere and neutral-sincere), $F(1, 174) = 13.10, p < .001$, simply reflecting that the mean F0 was higher overall. There was no difference between neutral-sincere and sweet-sincere utterances. Planned contrasts found that the utterances’ contour was significantly linear, $F(1, 174) = 11.99, p < .001$, but not quadratic. Figure 1 shows the overall intonation contour for the book pages.

**Discussion**

When mothers expressed humorous versus sincere communicative intentions, they used a higher mean F0, but did not differ in other F0 measures, or in intensity or speech rate. No differences were found between the two types of sincere communicative intentions, sweet-sincere and neutral-sincere. Mothers did not differentiate intonation contours for humorous versus sincere communicative intentions. Mothers used a linear, slightly rising contour overall. This could be due to the fact that the book was a counting book, and as such, this contour could reflect the list-like nature of the book. Selting (2007) reported that German speakers express open-ended lists with a rising or rising-plateau pitch, and tend to repeat the same intonation structure. Thus the fact that this was a counting book could have affected the intonation contour throughout the book, and as a result obscured the differences between communicative intentions.

**STUDY 2**

In Study 2, mothers’ use of vocal acoustic cues when expressing verbal humour and sincerity to their toddlers was investigated. Since toddlers come to appreciate verbal
humour later than visual humour (McGhee, 1979), it was predicted that mothers would increase their use of acoustic cues to help scaffold this type of humour. This study was necessary to determine whether mothers give additional cues for scaffolding a more difficult (verbal) form of humour. In this study, we compared humorous intentions against just one sincere intention, sweet-sincere. The content of the sentences was either humorous or sweet-sincere, and pictures served to highlight the sentence content. The sentences were also more complex, as they were not simply about counting, but described humorous or sweet-sincere concepts, such as having a cuddle (sweet-sincere), or a mother drinking from a baby’s bottle (humorous, see Appendix B). Mothers read either a book containing humorous sentences accompanied by humorous images, or a book containing sweet-sincere sentences accompanied by sweet-sincere images. The two books were edited versions of the book used in Study 1. The books were matched for pictures (containing the same general content, such as a baby, a mother, and a bottle), sentence structure, and general content, controlling for error due to any of these factors. More importantly, by matching the sentence structure, intonation patterns could be compared syllable by syllable. Furthermore, as the sentences were matched for length, this could reduce error for vocal cues such as speech rate. Finally, the edited books used in this study eliminated counting to avoid any list-like intonation patterns. Thus, Study 2 is a methodological improvement over Study 1. In Study 2, we also sought to determine whether toddlers respond differently to cues for humorous and sweet-sincere intentions. To do so, we measured toddlers’ smiling and visual attention at the end of each utterance, and the presence of toddlers’ laughter throughout the reading session. We chose only to compare humorous versus sweet-sincere utterances as sweet-sincere utterances are a better control than neutral-sincere utterances. Thus, we could focus on our primary research question - whether mothers give different acoustic cues for different types of communicative intentions (not emotions).

**Method**

**Participants**
Forty-six mother–toddler dyads participated. Five mother–toddler dyads were dropped from the study because the children were restless, leaving 41 mother–toddler dyads. Toddlers were aged 18 to 24 months with a mean age of 20 months and 9 days and a standard deviation of 1 month and 8 days. Twenty-seven of the toddlers were girls and 14 were boys. Twenty-two mother–toddler dyads read the humorous book and 19 mother–toddler dyads read the sweet-sincere book.

**Materials**
The books used in this study were edited versions of One Gorgeous Baby by Oborne and Godon (2004), which each contained six pages total. The book for the sweet-sincere condition was re-named One Lovely Baby and contained very sweet-sincere sentences for each page, for example, ‘Baby loves mummy’s cuddle’. The book for the humorous condition was re-named One Funny Baby and contained very humorous sentences for each page, for example, ‘Mummy drinks baby’s bottle’. Some pictures in the humorous condition were also altered to correspond to sentence content. Thus, humorous pages and sweet-sincere pages sometimes had different pictures (e.g., a mother cuddling a baby drinking a bottle [sweet-sincere] versus a mother drinking a baby’s bottle while
holding a baby [humorous]) but pictures contained the same elements (e.g., bottle, baby, and mother). There were also opening and closing pages for both books that were not analysed. In total, each version of the book contained four pages that were either humorous or sweet-sincere (see Appendix B). Twelve parents and adults who work with toddlers rated the book pages for humour and sweetness on a scale of 1–5. Paired-samples \( t \)-tests found that the humorous (\( M = 4.17, SD = 0.57 \)) pages were rated as significantly more humorous than the sweet-sincere pages (\( M = 1.58, SD = 0.59 \)), \( t(11) = 12.61, p < .001 \), and that the sweet-sincere pages (\( M = 3.58, SD = 0.78 \)) were rated as significantly more sweet than the humorous pages (\( M = 2.00, SD = 0.76 \)), \( t(11) = 4.73, p < .01 \). The equipment, software, and testing room were the same as in Study 1.

**Design**

This was a between-subjects experiment. The between-subjects independent variable was the communicative intention that the book pages conveyed, that is, humorous or sweet-sincere. The dependent variables included the same summary features as Study 1. The intonation-dependent variables included the shape of the intonation contours (linear, quadratic, cubic, or none of the above). For child responses, the dependent variables were whether the child smiled, and where the child looked (at the book, at the mother's face, or elsewhere) at the end of each page, and whether the child laughed during the book-reading session.

**Procedure**

The procedure was the same as in Study 1.

**Coding**

For each mother on each measure (e.g., \( F_0 \), intensity measures), the results for each of the humorous pages (Humour 1, Humour 2, Humour 3, and Humour 4) were averaged out to give a score for humorous utterances. The same was done for sweet-sincere utterances (Sweet 1, Sweet 2, Sweet 3, and Sweet 4). Where a mother had not said one or more of the sentences within a category or one or more of the utterances had been masked by noise (emanating from the child), whichever utterances were present in a category for a mother were averaged.

One participant in the sweet-sincere condition was dropped from the analyses of \( F_0 \) and intensity variables, but not speech rate, because all of her utterances contained noise emanating from the child. Another participant in the humorous condition was dropped from intensity analyses, but not \( F_0 \) or speech rate analyses, because she sat on the floor away from the microphone rather than on the sofa next to it.

For each mother, the four utterances were re-digitized from DV video recording to wav files as in Study 1. Mean, range, and standard deviation of \( F_0 \), and mean, range, and standard deviation of intensity and speech rate were coded as in Study 1.

Utterances Humour 1, Humour 2, Humour 3, Sweet 1, Sweet 2, and Sweet 3 were examined for intonation contours. For each utterance, the coder identified syllable boundaries by visually and auditorily inspecting the contour. For each syllable, the mean \( F_0 \) was calculated by averaging the \( F_0 \) values across all time steps in that syllable in semitones. Multiple utterances (e.g., Humour 1, 2, and 3) were not averaged
Table 2. Means, standard deviations, and independent-samples t-tests for vocal summary features between humorous and sweet-sincere book pages in Study 2

<table>
<thead>
<tr>
<th>Summary feature</th>
<th>Humorous</th>
<th>Sweet-sincere</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F0 (st)</td>
<td>16.15</td>
<td>12.62</td>
<td>6.52</td>
<td>38</td>
<td>&lt;.001</td>
<td>2.24</td>
</tr>
<tr>
<td>SD</td>
<td>(1.44)</td>
<td>(1.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 standard deviation (st)</td>
<td>4.83</td>
<td>3.72</td>
<td>3.16</td>
<td>38</td>
<td>&lt;.01</td>
<td>1.01</td>
</tr>
<tr>
<td>SD</td>
<td>(1.09)</td>
<td>(1.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 range (st)</td>
<td>19.65</td>
<td>15.61</td>
<td>3.16</td>
<td>38</td>
<td>&lt;.01</td>
<td>1.01</td>
</tr>
<tr>
<td>SD</td>
<td>(3.45)</td>
<td>(4.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intensity (dB)</td>
<td>45.24</td>
<td>44.49</td>
<td>2.35</td>
<td>34.95</td>
<td>&lt;.05</td>
<td>0.75</td>
</tr>
<tr>
<td>SD</td>
<td>(1.12)</td>
<td>(0.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity standard deviation (dB)</td>
<td>1.53</td>
<td>1.02</td>
<td>2.11</td>
<td>37</td>
<td>&lt;.05</td>
<td>0.70</td>
</tr>
<tr>
<td>SD</td>
<td>(0.85)</td>
<td>(0.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity range (dB)</td>
<td>6.72</td>
<td>4.62</td>
<td>2.11</td>
<td>37</td>
<td>&lt;.05</td>
<td>0.71</td>
</tr>
<tr>
<td>SD</td>
<td>(3.35)</td>
<td>(2.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech rate (syllables/second)</td>
<td>2.89</td>
<td>3.45</td>
<td>3.43</td>
<td>39</td>
<td>&lt;.01</td>
<td>1.10</td>
</tr>
<tr>
<td>SD</td>
<td>(0.42)</td>
<td>(0.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

for intonation contour as each utterance contained a different number of syllables. Humour 4 and Sweet 4 were not included because the sweet and humorous versions contained different numbers of syllables. Twenty percent of the utterances were coded by a second coder, and the correlation between mean F0 for each syllable was \( r = 0.98 \), \( n = 279 \), \( p < .01 \).

Children’s responses were coded for (1) smiling at the end of each coded utterance; (2) visual attention at the end of each coded utterance (to book, mother’s face, or elsewhere); and (3) laughter during the book-reading session. Due to technical problems, one video was not coded. Twenty-five percent of the videos were coded by a second coder. Reliability was 97% for smiling, 89% for visual attention, and 100% for laughter.

Results

Mothers’ head position in relation to the microphone was coded for each utterance in case this affected the results of the acoustic analyses. Three percent of mothers’ heads were turned at 180° from the microphone; 81% were turned at 135°; and 16% were turned at 90°. Independent-samples t-tests found no differences between head positions in the humorous and sweet-sincere pages.

Preliminary analyses found no effects of child gender or child age (under or over the mean age of 20 months) on any of the variables. Thus, child gender and age were dropped from all analyses.

Summary features

The means, standard deviations, and independent-samples t-tests for each of the summary features (mean F0, F0 standard deviation, F0 range, mean intensity, intensity standard deviation, intensity range, and speech rate) can be found in Table 2. When mothers expressed humour as compared to sweet-sincerity, they used a significantly higher mean F0, a significantly higher F0 standard deviation, and a significantly larger F0 range. When
testing mean intensity, Levene’s test of equality of variance was violated, $F = 4.41, p < .05$. When equal variance was not assumed, mothers expressing humour as compared to sweet-sincerity used a significantly higher mean intensity. Mothers expressing humour as compared to sweet-sincerity also used a significantly higher intensity standard deviation, a significantly wider intensity range, and a significantly slower speech rate.

**Intonation contours**
The mean $F_0$s for each syllable were entered into Mixed Models in order to discover whether the utterance had a linear, quadratic, or cubic (wave shaped; rise-fall-rise or fall-rise-fall) intonation contour, based on the intonation contours from Katz *et al.* (1996). Furthermore, as the sentences were matched for sentence structure in the two books, humorous and sweet-sincere communicative intentions were compared to see if the intonation contours were the same or different. Figure 2 shows the intonation contours for utterances Humour 1, Humour 2, and Humour 3 in the humorous book, and Sweet 1, Sweet 2, and Sweet 3 in the sweet-sincere book.

A 2 (communicative intention: humorous, sweet-sincere) × 7 (syllable: mean $F_0$ of each syllable) Mixed Model for Humour 1 and Sweet 1 combined, a 2 (communicative intention) × 13 (syllable) Mixed Model for Humour 2 and Sweet 2 combined, and a 2 (communicative intention) × 7 (syllable) Mixed Model for Humour 3 and Sweet 3 combined were run. Effects of communicative intention by syllable were found for all three pages (Humour 1 and Sweet 1: $F(6, 270) = 5.32, p < .001$; Humour 2 and Sweet 2: $F(12, 465) = 2.63, p < .01$; Humour 3 and Sweet 3: $F(6, 265) = 13.20, p < .001$). Effects of communicative intention were found, (Humour 1 and Sweet 1: $F(1, 270) = 116.05, p < .001$; Humour 2 and Sweet 2: $F(1, 465) = 44.23, p < .001$; Humour 3 and Sweet 3: $F(1, 265) = 101.68, p < .001$) simply reflecting the fact that humorous utterances were higher in pitch overall. Additionally effects of syllable were found (Humour 1 and Sweet 1: $F(6, 145) = 5.92, p < .001$, $Pseudo-R^2 = .197$; Humour 2 and Sweet 2: $F(12, 256) = 2.99, p < .01$, $Pseudo-R^2 = .123$; Humour 3 and Sweet 3: $F(6, 147) = 18.52, p < .001$, $Pseudo-R^2 = .430$). Planned contrasts found that all three utterances’ intonation contours were significantly linear, (Humour 1: $F(1, 145) = 21.11, p < .001$; Humour 2: $F(1, 256) = 6.74, p < .025$; Humour 3: $F(1, 147) = 65.09, p < .001$). Humour 1 and Humour 3, but not Humour 2, also showed significant quadratic trends (Humour 1: $F(1, 145) = 6.29, p < .025$; Humour 3: $F(1, 147) = 11.82, p < .01$). However, linear contours displayed the best fits.

A 7-way (syllable), a 13-way, and a 7-way Mixed Models for Sweet 1, Sweet 2, and Sweet 3, respectively, were run. An effect of syllable was marginally significant for utterance Sweet 2, $F(12, 209) = 1.79, p = .051$, $Pseudo-R^2 = .093$. A cubic, but neither a linear nor quadratic, effect was found for the intonation contour of Sweet 2, $F(1, 209) = 5.47, p < .025$. There were no effects of syllable for Sweet 1 or Sweet 3.

**Children’s responses**
Since children smiled and looked away from the book to the mother’s face around one time on average or less across all four pages in both humorous (smile: $M = 0.36$,
Figure 2. Intonation contours for utterances Humour 1 and Sweet 1, Humour 2 and Sweet 2, and Humour 3 and Sweet 3, respectively, in Study 2.
SD = 0.79; look to parent: M = 1.00, SD = 1.35) and sweet-sincere (smile: M = 0.22, SD = 0.73; look to parent: M = 1.04, SD = 1.23) conditions, we used non-parametric statistics. Mann–Whitney U-tests revealed that toddlers were significantly more likely to laugh when read the humorous (8/22) versus the sweet-sincere book (1/18), U = 144.00, p < .025. Toddlers were equally likely to smile when read humorous (5/22) versus sweet-sincere (2/18) pages, and equally likely to look to the parent when read humorous (9/22) versus sweet-sincere (9/18) pages. A bi-variate correlation found that laughter was significantly correlated to mean F0, \( r = .339, p < .05 \). Bi-variate correlations found no relationship between the acoustic summary features of each utterance and either smiling or looking behaviour.

**Discussion**

The results of Study 2 demonstrate that mothers make use of a variety of vocal cues to differentiate between humorous and sincere communicative intentions. As in Study 1, mothers expressing humour used a higher mean F0 compared to mothers expressing sweet-sincerity. In addition, mothers expressing humour in Study 2 also used a higher F0 standard deviation, a wider F0 range, a higher mean intensity, a wider intensity range, a higher intensity standard deviation, and a slower speech rate compared to mothers expressing sweet-sincerity. Furthermore, in Study 2, mothers expressing humour also used significantly different intonation contours for all three humorous/sweet-sincere utterances measured (Humour 1 vs. Sweet 1; Humour 2 vs. Sweet 2; Humour 3 vs. Sweet 3). The best fit for all three humorous utterances was a rising linear contour, while sweet-sincere utterances generally did not follow a significant pattern, with Sweet 2 showing only a marginally significant cubic trend.

While children’s smiling and looking behaviours did not differ in relation to the communicative intentions of the book pages or to any acoustic features of mothers’ speech, children’s laughter was significantly more prevalent in the humorous book. Children’s laughter was also significantly correlated to mean F0.

While head position did not vary significantly between conditions, it is still possible that differences in head position, distance, and height added uncontrolled variability to the intensity measures. Future studies should use microphones that are mounted at a constant position, distance, and height away from the mothers’ mouths (e.g., a head-mounted microphone held in place a set distance from the mothers’ mouths).

**GENERAL DISCUSSION**

This study investigated whether people use unique vocal acoustic cues to differentiate positive communicative intentions. We investigated this by asking mothers to read to their toddlers, and by analysing the acoustic features of reading-based speech. Study 1 showed that mothers use a higher mean F0 to cue their toddlers to visual humour (e.g., a picture of a child throwing nappies/diapers) as compared to sincere images (sweet-sincere, e.g., a picture of a child holding and playing with toys; and neutral-sincere, e.g., a picture of a child dressed in a coat to go outside). Thus, mothers do differentiate between positive communicative intentions, in particular cuing their toddlers to visual humour by varying mean F0. Mothers did not distinguish sweet-sincere and neutral-sincere utterances.

In Study 2, the sentence content was humorous in and of itself (e.g., ‘Mummy drinks baby’s bottle’), or sweet-sincere in and of itself (e.g., ‘Baby loves mummy’s cuddle’),
heightened by humorous or sweet-sincere pictures. Study 2 showed that mothers use a higher mean F0, larger F0 standard deviation, wider F0 range, larger intensity standard deviation, wider intensity range, and slower speech rate when expressing verbal humour versus verbal sweet-sincerity. The second study also showed that humorous and sweet-sincere utterance contours were significantly different. Humorous utterances followed a rising linear contour, whereas sweet-sincere utterances did not tend to follow any specific pattern.

An important difference between Study 1 and Study 2 was the modality of humour: Study 1 contained visual humour only, in that the sentences were not humorous in and of themselves, but the pictures were humorous. In contrast, Study 2 contained verbal humour, such that sentences were humorous in and of themselves, heightened by visual images. Thus, mothers use more precise vocal acoustic cues, including intonation contour, to differentiate verbal humour and sincerity. McGhee (1979) suggested that toddlers appreciate visual humour from 18 months, and verbal humour from 24 months. Since the children in these two studies fall between the ages of 18 and 24 months, this suggests that mothers reserve certain cues for scaffolding verbal humour since it is more difficult for 18- to 24-month-olds to appreciate than visual humour.

The results of these studies demonstrate that mothers use specific acoustic cues when expressing humorous versus sincere (sweet-sincere and neutral-sincere) communicative intentions. This could help toddlers come to understand that their mothers are not being serious when saying things like ‘Mummy drinks baby’s bottle’, but instead intend the utterance to be a joke. This challenges the idea that understanding mental states occurs strictly through children’s innate capacity to do so (Leslie, 1987). Instead, our data suggest that acoustic cues are available to allow children to differentiate communicative intentions, suggesting that social input could help toddlers in this task.

**Toddlers’ responses**

Toddlers were more likely to laugh if being read the humorous book, and when mothers used a higher mean F0. As mothers used a higher mean F0 when reading humorous book content, it is difficult to determine whether toddlers were laughing at humorous book content only, heightened mean F0 only, or a combination of the two. Experimental research should isolate which of these factors elicit toddlers’ laughter.

There was no difference in smiling or looking behaviours between book and page types in Study 2. Furthermore, there was no relationship between the toddlers’ smiling or looking behaviours and any of the acoustic measures. This suggests that toddlers do not react automatically to intonation or utterances. Since toddlers may need to (1) realize that the utterance should be attended to; (2) realize that the utterance should not be taken seriously; and (3) figure out what exactly made the joke humorous, it may be the case that toddlers would show more smiling and looking behaviours after a delay of several seconds, rather than immediately as we measured. Future experiments should examine children’s delayed responses to humorous, neutral-sincere, and sweet-sincere intonation patterns and book pages.

Hoicka and Wang (2011) found that when 15-month-olds heard humorous vocal acoustic cues (laughter and an utterance that was louder, higher, and had a rising contour), they expected to see a humorous versus sweet-sincere action. Conversely, when 15-month-olds heard sweet-sincere vocal acoustic cues (the expression, ‘Awww!’)
and an utterance that was quieter, lower, and had a rise-fall contour to stand in as a non-pragmatic contour, following Katz et al., 1996), they expected to see a sweet-sincere versus humorous action. This suggests that infants do process acoustic information pertaining to humorous and sweet-sincere cues in a meaningful way.

**Humour and pretend play**

The results of Study 1 complement Reissland and Snow’s (1996) report that mothers increase their mean F0 during a pretend-play context. This similarity may reflect the expression of elation in both contexts, which is characterized in part by an increased mean F0 (Scherer, 1986; Banse & Scherer, 1996). However in Study 1, the acoustic features of humour were not only compared to a neutral-sincere control, but also to a sweet-sincere control, suggesting that elation was not the basis for the acoustic profile of humour. The similarity between humour and pretend play in acoustic profiles could however be due to similar communicative intentions, since humour and pretend play both convey abstract concepts (McGhee, 1979; Reddy, 1991, 2001) and the intention to do the wrong thing (Leekam, 1991; Hoicka & Gattis, 2008). Future studies might compare acoustic features of humour and pretend play to clarify this further.

**Why is humour louder and higher, but also slower?**

The acoustic features of verbal humour can be compared to several previously studied acoustic patterns. A higher mean F0 may be a result of smiling (Ohala, 1984, 1994) which would be more likely during the expression of humour over sincerity. Additionally, Scherer (1986) and Banse and Scherer (1996) found that an elated or joyous expression is higher in mean F0, F0 range, F0 variability, mean energy (intensity), energy range, energy variability, and speech rate compared to a neutral standpoint. Therefore, it is not the case that the expression of humorous communicative intentions is simply the expression of elation in Study 2 because of one factor: the speech rate in humour is slower than sweet-sincerity. Thus, humorous vocal acoustic signals are distinct from elated acoustic signals.

The combined acoustic features of higher, louder, and slower are also characteristic of IDS. Fernald and Simon (1984) found that IDS is characterized by an increased mean F0. Beckford Wassink, Wright, and Franklin (2007) found that IDS is louder. Finally, Papoušek, Papoušek, and Haekel (1987) found that IDS is characterized by a slower speech rate. It is proposed that humour is louder, higher, and slower because mothers use these vocal features in order to bid for the toddler’s attention and make the utterance content easier to understand. This interpretation is based on three previous studies. First, Fernald and Simon (1984) suggested that increased pitch used towards infants could indicate an attention bid. Second, Newman (2003) reported that in noisy (vs. quiet) situations, parents use aspects of IDS in order to teach their children a new word, including higher mean F0 and slower speech rate. Newman suggested that this is to make it easier for the toddler to segment what the parents are saying, and hence better understand the utterance. Since toddlers are thought to understand verbal humour only at the end of the second year (McGhee, 1979), it is suitable that mothers in this study are using their voices to effectively teach verbal humour appreciation.
Rising linear contour

The influence of communicative intentions on intonation contour indicates that IDS cannot, however, fully explain the results. In Study 2, humorous utterances were rising and linear, whereas no pattern was found in the intonation contours for sweet-sincere utterances. The features of humorous intonation contours are not global features of IDS. Indeed, intonation contours map onto pragmatic or linguistic categories (e.g., Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985). We propose that this feature is used by mothers to let their children know that what they are saying is open to interpretation, is questionable, and is not a fact, or a declarative.

Whereas statements are expressed using falling contours, rising contours are associated with ‘testing’, expressing an utterance as an interrogative or question and inviting the listener to resolve the issue (Ohala, 1984, 1994; van Heuven & Haan, 2000; Gussenhoven, 2004). Thus, mothers can use a rising linear intonation contour to indicate that a joke has questionable content, and should thus not be interpreted as new, true, information. Furthermore, Ladd et al. (1985) found that adults judge rising contours as signalling contradiction, and Goodwin, Goodwin, and Yaeger-Dror (2002) found that girls used pitch prominence (increased pitch) when expressing opposition or negation during games. This fits well with the study by Hoicka et al. (2008) in which parents made disbelief statements after making jokes. Like disbelief statements, a rising linear contour could be another cue to let toddlers know that the utterance is not intended to be believed.

Based on these reports, we propose that mothers communicating verbal humour use an extreme form of IDS to elicit toddlers’ attention and make it easier for them to understand the content of the joke. In addition, mothers communicating verbal humour use a rising linear contour to indicate that the utterance is open to interpretation and is not in fact a statement or declarative. With these acoustic features, mothers can signal to their children that what they have said is not meant to be taken literally (as in the case of sweet-sincerity), but is instead meant to be a joke, and invite their children to resolve the incongruity.

Future directions

Future research should examine whether these cues are humour-specific, or more general to intentionally saying the wrong thing. In particular, humour and pretend play should be directly compared. A fuller range of vocal acoustic cues for pretend play could be measured to compare to humour. Furthermore, a controlled non-pretend play condition could be introduced to ascertain whether the cues are due to play (and thus more likely elation) or are more specific to pretend play (and then perhaps abstraction or incongruity).

Lehiste and Peterson (1959) found that different vowels have different intrinsic intensities. While a basic look at the vowels in both studies do not suggest that the humorous conditions contained intrinsically louder vowels, future studies would ideally match sentences exactly in the humorous and non-humorous conditions in order to be certain that intensity (and other) effects were due to the communicative intention alone.

Finally, while we only studied mothers, we expect that the results should also extend to fathers. For example, fathers, like mothers, use features of child-directed speech (e.g., Fernald et al., 1989; Shute & Wheldall, 1999), although some features may be different to those typically used by mothers. Future research should examine how fathers express humour to toddlers.
Conclusions
The studies in this article have shown that mothers use distinct vocal cues to differentiate humorous and sincere communicative intentions. We suggest that mothers use extreme characteristics of IDS when expressing humour (versus sincerity) to focus toddlers' attention and make it easy for them to understand the utterance being said. We also suggest that mothers use a rising linear contour when expressing humour to indicate that the utterance is not a statement, but is open to interpretation. Thus toddlers' acoustic environments contain rich cues denoting humorous versus sincere communicative intentions.

Acknowledgements
The research reported here was supported by an Overseas Research Scholarship from the Overseas Research Student Awards Scheme, a Research Studentship from Cardiff University, a Margaret Addison scholarship from the Victoria Women's Association, University of Toronto, and a post-doctoral research fellowship from the National Institutes of Health (T32-HD046423) all awarded to Elena Hoicka. We thank all the parents and infants who participated. We also thank Sarah Jutsum for her assistance with testing, Jack Vevea for statistical help, and Nadja Reissland for stimulating discussions about vocal cues to mental states.

References


Received 10 January 2011; revised version received 1 September 2011