Inexpensive video cameras used by parents to record social communication in epidemiological investigations in early childhood—A feasibility study

Philip Wilsona,⁎, Christine Puckeringb, Alex McConnachiec, Helen Marwickd, Nadja Reisslande, Christopher Gillberga

a Community Based Sciences, Faculty of Medicine, University of Glasgow, United Kingdom
b Royal Hospital for Sick Children, Glasgow, United Kingdom
c Robertson Centre for Biostatistics, University of Glasgow, United Kingdom
d National Centre for Autism Studies, Strathclyde University, United Kingdom
e Department of Psychology, Durham University, United Kingdom

Article history:
Received 31 October 2009
Received in revised form 8 June 2010
Accepted 24 September 2010

Keywords:
Social interaction
Video camera
Communication
Autism
Infants

A B S T R A C T

We tested the feasibility of parents recording social interactions with their infants using inexpensive camcorders, as a potential method of effective, convenient, and economical large scale data gathering on social communication. Participants were asked to record two short video clips during either play or a mealtime, and return the data. Sixty-five video clips (32 pairs) were returned by 33 families, comprising 8.5% of families contacted, 44.6% of respondents and 51.6% of those sent a camcorder, and the general visual and sound quality of the data was assessed.

Audio and video quality were adequate for analysis in 85% of clips and several social behaviours, including social engagement and contingent responsiveness, could be assessed in 97% of clips. We examined two quantifiable social behaviours quantitatively in both adults and infants: gaze direction and duration, and vocalization occurrence and duration. It proved difficult for most observers to obtain a simultaneous clear view of the parents and infant’s face.

Video clips obtained by parents are informative and usable for analysis. Further work is required to establish the acceptability of this technique in longitudinal studies of child development and to maximize the return of usable data.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

There is a significant body of evidence that predictors of serious psychopathology, such as autism and attention-deficit/hyperactivity disorder (ADHD) exist in early childhood (Trevarthen, 2001) but recall bias, small sample sizes and other methodological considerations limit their general predictive and clinical value (Zwaigenbaum et al., 2005) Attempts

⁎ The research described in this paper was conducted in accordance with APA ethical standards in the treatment of the study sample.

‡ The study was approved by Greater Glasgow Primary Division Local Research Ethics Committee (reference (PCD2-GP) 05/50706/63).

⁎⁎ Corresponding author at: General Practice and Primary Care, 1 Horselethill Rd, Glasgow G12 9LX, United Kingdom. Tel.: +44 141 632 6310; fax: +44 141 636 1180.

E-mail address: p.wilson@clinmed.gla.ac.uk (P. Wilson).

0163-6383/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.infbeh.2010.09.007
have been made to develop clinical tools for early identification of vulnerabilities, such as the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen, Allen, & Gillberg, 1992) and the M-CHAT (Robinson, Puckering, Wright, & Parkinson, 2009) and these could, with modifications to increase sensitivity and specificity, potentially be of value in identifying children for earlier appropriate interventions than is currently possible. Population screening methods for autism, ADHD and other early onset psychopathology need to be shown to be economical, sensitive and specific in identifying early interactional behaviours that differentiate normal from deviant development. Measures obtained from young children by professional observers are promising but potentially time consuming and a search for more efficient methods that can be shown to predict later pathology is needed before effective early intervention strategies can be developed (Wilson et al., 2009). Constructs such as interpersonal synchrony (Papousek & Papousek, 1997), mutuality, support, positive attitude, infant centred speech (Murray, Kempton, Woolgar, & Hooper, 1993) and attachment (Bor, Brennan, Williams, Najman, & O’Callaghan, 2003) have been generated as predictors of psychopathology but need to be applicable in a simplified format if they are to be used clinically or with larger samples in epidemiological studies.

The value of parent–child interaction recorded in the home lies in its ability to capture the quality and style of this interaction as well as quantifiable factors such as infant and parent activity and attention levels. The presence of a camera can be expected to influence the interaction and behaviour to some extent (Penner et al., 2007), although a self-operated system may be expected to reduce the effects of the presence of an observer (Cohen, Manion, & Morrison, 2007).

Context of recording, such as feeding or object play, has been found to influence interactive engagement (Leyendecker, Lamb, & Schölmerich, 1997). A mealtime may be an ideal interaction for recording. It brings parent and child into proximity, involves the negotiation of mess, self-feeding, amount and type of food to be consumed, and the agreed end of the meal. The familiarity of the situation reduces self-conscious behaviours and elicits characteristic patterns of interaction. The play setting is more variable and may be less likely to elicit habitual behaviour than normal daily caretaking, but it does permit assessment of the age-appropriateness and ‘child-centredness’ of the type of play (Crittenden, 1988).

Video recording of parent–child interactions has become a standard tool in the clinical assessment of social communication and its disorders (Crittenden & Bonvillian, 1984; Gunning et al., 2004). The fine-grained analysis of parent–child interaction has been the subject of intense academic interest (Gunning, Fiori-Cowley, & Murray, 1999; Murray, 1992; Trevarthen, 1979; Trevarthen, 2001) and many authors have identified aspects of the interaction which correlate strongly with psychopathology in the mother (Crittenden & Bonvillian, 1984; Hipwell, Goossens, Melhuish, & Kumar, 2000; Reissland, Shepherd, & Herrera, 2003; Trevarthen, 2001), particularly postnatal depression (Field, Sandberg, Vega-Lahr, Goldstein, & Guy, 1985; Field et al., 1988; Field, Healy, Goldstein, & Guthertz, 1990; Field, 1995) and later in the child (Maestro, Casella, Milone, Muratori, & Palacio-Espa, 1999; Maestro et al., 2001; Murray, Fiori-Cowley, Hooper, & Cooper, 1996; Murray, Woolgar, Cooper, & Hipwell, 2001; Murray, Marwick, & Arteche, 2010; Osterling & Dawson, 1994; Osterling, Dawson, & Munson, 2002; Werner & Dawson, 2005; Werner, Dawson, Osterling, & Dinno, 2000). Some of these studies (Maestro et al., 1999; Maestro et al., 2001; Osterling & Dawson, 1994; Osterling et al., 2002; Werner et al., 2000; Werner & Dawson, 2005) are noteworthy because marked abnormalities of communication style were noted in home video material captured during infancy among children with autism, long before formal diagnosis. There are however no detailed representative community-based studies of parent–child interaction on a large enough scale to be useful in setting normative thresholds or identifying early markers of psychopathology. The present study was designed with a view to testing a new and relatively simple methodology for the objective study of real-life parent–child interaction, with the further goal of developing better instruments for screening forerunners of early onset psychopathology, such as autism, ADHD and conduct disorder.

The price of video recording has plummeted in recent years. Several commercially available video cameras capable of recording a few minutes of reasonable quality video are available for under £40 – and prices continue to fall. Most modern mobile telephones are now capable of recording video footage. It is therefore now potentially possible to record human interactions in large studies designed to investigate social communication using technology which is cheap, accessible and relatively familiar and non-threatening to families.

2. Methods

Families with infants aged 8–11 months were identified in participating National Health Service general medical practices in Glasgow, Scotland, serving a range of socio-economic environments. The list of infants was scrutinised by one or more clinicians in each practice who could use their discretion to exclude families. An information leaflet was posted to eligible families, along with two questionnaires regarding the approach: one for those accepting the invitation to participate, one for those declining.

Families responding positively were sent a video camcorder (Mustek DV 5200) with a 20 cm high tripod, two 64 MB memory cards (capable of holding a recording of at least three minutes duration), usage instructions, a covering letter, a further consent form, a questionnaire (eliciting their views on the approach, any problems they had with the recording and other aspects of the experience for the family) and a reply-paid envelope. It became apparent three months into the study that families often took many weeks to return the video clips, so we started to request the return of the video material within a week of receipt of the camera.

Participants were assigned at random to a request to record either (i) three minutes of a normal mealtime for the child, with both principal carer and child in the frame on two separate occasions, (ii) three minutes of normal play on two separate occasions with the principal carer or (iii) three minutes of a meal and three minutes of play. These conditions were selected
to allow us to compare the utility of the two settings for eliciting social behaviours, and to examine the consistency of these behaviours both across conditions and across repeated episodes of the same condition.

Families returning the memory cards had been informed that the research team would return them together with a VHS tape or DVD of the video sequence, and a letter was sent on receipt of the clips stating that the family could keep the camcorder.

2.1. Analysis

Video cards were received and the data transferred to computer on receipt at the study office. Video file formats were converted (into .mpg format) for detailed behavioural analysis, and separate sound files (.wav) were created from the video clips in order to examine audio data using Praat spectrographic software (www.fon.hum.uva.nl/praat/). Three-minute clips were prepared from these data, starting at the first point at which both parent and child were visible. These three-minute clips were further divided into three one-minute segments.

Video segments were analysed using Noldus Observer XT software version 8.0. Audio data were processed using Praat freeware, and the pitch data imported into the Observer software using the external data module (details available from the authors on request).

2.2. Measures

The following were assessed in the three one-minute segments of each video sequence:

- Whether a substantial part of the parent’s face could be seen. Judged by rating whether two eyes, one eye, or neither eye were visible the majority of the time.
- Whether a substantial part of the infant’s face could be seen. Judged as above.
- Whether the sound was of sufficient quality to distinguish adult and infant vocalisations and whether further analysis of vocal quality (eg pitch) could be performed on both infant and adult vocalisations.
- Whether the parent or infant exhibited behaviours contingent upon the actions or situation of the other, i.e. turn-taking. Contingent responsiveness was a subjective judgement of the timing and connection between the social behaviours of the adult and infant – such as the infant turning to look at the mother following the mother calling the infant’s name. The purpose of this analysis was to establish that the quality of the video allowed contingent timing and connections between social behaviours to be identified, rather than to generate a detailed analysis of their occurrence and patterning.
- Whether the parent positively engaged either physically or verbally with the infant and whether the infant positively engaged either physically or verbally with the parent. This was again a subjective observational analysis using a broad definition of interpersonal behaviour, which comprised any behaviours and expressiveness that seemed directed towards or responsive to the other person, leading to a judgement of interpersonal engagement and involvement. This would include behaviours such as looking at, smiling at, reaching out towards, leaning towards, and vocalisations, which could occur singly or in combinations. The purpose of the analysis was again to establish that the quality of the video allowed such behaviours to be identified, rather than to generate data on their occurrence and patterning.

In addition, the following variables were assessed in all of the video clips that were of sufficient quality:

- Duration of episodes of gaze at the face of the other person between infant and parent.
- Duration of infant’s and parent’s vocalisations.

A second observer also rated the above variables in 10% of the videos, which were also re-rated after an interval (of at least two weeks) by the first observer.

2.3. Statistical analysis

Amongst those families who responded to the invitation to take part, the deprivation category (Carstairs & Morris, 1991; McLoone, 1994), age and age left full time education of the respondent were compared between those that declined to participate, agreed but failed to return any video material, or agreed and returned the video clips. Fisher’s Exact Test and Kruskal–Wallis Tests were used.

Basic quality assessments of each video clip are reported in terms of the number of one-minute segments (0–3) that met each quality criterion.

In a sample of 20 separate one-minute clips, assessments of gaze and vocalisation were made by two raters and by one rater on two occasions; inter- and intra-rater reliability are summarised in terms of the intraclass correlation coefficient (ICC, estimated using variance components models), and the Spearman Rank correlation coefficient. When preliminary analyses suggested systematic discrepancies between the two raters in relation to their assessments of infant vocalisation, two additional raters were recruited to provide additional data for this variable.
Similar methods were applied to estimate the reliability of assessments made by a single rater between one-minute segments within the same video, and between videos of the same child, the latter are reported as a whole and separately for those families asked to provide either two videos of a meal setting, two of a play setting, or one of each.

3. Results

Four hundred and nineteen families with infants aged 8–11 months were identified in 25 practices. Thirty-one (7.4%) families were excluded by clinicians. Reasons for exclusions included language barriers, illness, social problems and non-attendance. No reason was given for the exclusion of eight of the 31 families.

An information sheet was posted to the 388 eligible families of which 74 (19.1%) responded. Sixty four (16.5% of families contacted, 86.5% of respondents) accepted the offer to record video footage with their infant. Of the 10 families declining the offer, reasons cited were time constraints (7), worries over the security of the video footage (4), male researchers (2) or strangers (4) in the study team and being camera shy (2). One of the families declining to participate said they would have accepted the offer if a researcher recorded the footage.

After the return of the video cards, of the 64 families who received camcorders, 28 (43.8%) preferred to make the recording themselves, 25 (39.1%) would have preferred a researcher to do it and 11 (17.2%) had no preference. Thirty-three (8.5% of families contacted, 44.6% of respondents and 51.6% of those sent a camcorder) returned video records of parent–child interactions.

Table 1 details the socioeconomic status, ages and educational level of respondents who declined to take part, agreed but did not return video material, or supplied video clips. There was little evidence of differences between the 3 groups of respondents, with the possible exception that those who responded to the invitation but declined to take part had left education at a younger age (p = 0.070).

Of the 64 respondents agreeing to participate, 55 (85.9%) reported discussing the research with their partners. The research was also discussed with parents (12), other relations (10) and friends (3); seven respondents reported not discussing the research with anyone. Of the 10 families declining to participate, partners (8), relatives (2) and friends (2) were also consulted. None reported discussing the research with their parents. There was however no statistical evidence that participants and non-participants differed in terms of sources of advice (Fisher’s Exact test, p = 0.20). Two non-participating respondents did not discuss the research with anyone.

Twenty-six (78%) of the 33 families returning clips found the instruction for recording footage ‘easy’ to follow, with the remaining 6 (18%) reporting the instructions as ‘OK’ (1 missing response). None of the families reported finding the instructions difficult, but 16 families (48%) failed to get the recording right first time around for at least one of the two recordings they were asked to provide, though only 3 (9%) needed more than one attempt on both occasions.

Reasons for difficulty in recording videos included uncertainty as to the sound and lighting quality or other technical problems (6), difficulty in positioning the camera correctly (5) and distractions for the baby (including the camera itself) (5). Eight participants noted that the behaviour of the child was not “usual” or “normal” in at least one of the recordings.

The 33 participating families were asked to provide two video recordings of a feed and/or a play situation; 65 video clips were returned (one family provided only one clip). Clips varied between two and 14 minutes in length. An initial inspection determined 64 clips (98.5%) to be of adequate sound quality for analysis. Each video clip was cut into three one-minute segments (where there was sufficient recording time), providing a total of 194 segments for closer analysis (one clip provided only two segments). Each segment was assessed for sound quality both in terms of being able to measure pitch and to distinguish between the baby’s and the parent’s voices (in order to measure vocalisation). Whether one or both of the baby’s and the parent’s eyes were visible for the majority of the segment was also assessed, in order to be able to measure gaze and as a proxy for general facial visibility. Table 2 summarises these assessments.

Sound quality was degraded in several segments by noisy toys, television or radio, or other family members, but in most cases was of adequate quality to assess pitch and to distinguish between parental and child vocalisations. Similarly, a full view of the baby’s face was seen on most segments, and a partial view on almost all. Whilst a full view of the parent’s face was not the norm in most one-minute segments, in the majority of cases, there was at least a partial view for most of the segment. Deprivation category, age of parent, age at completion of education, and play vs. feed condition were not found to impact on the proportion of clips with adequate video or sound quality.

Table 1
Responding families. Carstairs deprivation category (McLoone, 1994), where (1 = most affluent, 7 = most deprived), and respondent’s age and age left full time education.

<table>
<thead>
<tr>
<th>N</th>
<th>Deprivation category</th>
<th>Age or respondent</th>
<th>Age left FT education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1–2</td>
<td>4–5</td>
<td>6–7</td>
</tr>
<tr>
<td>Declined</td>
<td>10</td>
<td>3 (30.0%)</td>
<td>5 (50.0%)</td>
</tr>
<tr>
<td>Agreed, but no video returned</td>
<td>31</td>
<td>4 (12.9%)</td>
<td>15 (48.4%)</td>
</tr>
<tr>
<td>Agreed and returned video</td>
<td>33</td>
<td>7 (21.2%)</td>
<td>13 (39.4%)</td>
</tr>
<tr>
<td>p-Values</td>
<td>0.628</td>
<td>0.254</td>
<td>0.070</td>
</tr>
</tbody>
</table>

---

34 (2011) 63–71

Table 2
Adequacy of basic audio and video data. Number and percentage of videos providing 0, 1, 2 or 3 one-minute segments of adequate quality according to various criteria.

<table>
<thead>
<tr>
<th>Quality criterion</th>
<th>Number of one-minute segments meeting quality criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Sound differentiability (Parent vs. Child)</td>
<td>7 (10.8%)</td>
</tr>
<tr>
<td>Sound quality (Pitch Analysis)</td>
<td>2 (3.1%)</td>
</tr>
<tr>
<td>Full view of child's face</td>
<td>21 (32.3%)</td>
</tr>
<tr>
<td>At least partial view of child's face</td>
<td>3 (4.6%)</td>
</tr>
<tr>
<td>Full view of adult's face</td>
<td>49 (75.4%)</td>
</tr>
<tr>
<td>At least partial view of adult's face</td>
<td>10 (15.4%)</td>
</tr>
</tbody>
</table>

Table 3
Adequacy of data for analysis of social behaviours. Number and percentage of videos providing 0, 1, 2 or 3 one-minute segments where social behaviours were observed.

<table>
<thead>
<tr>
<th>Social behaviour</th>
<th>Number of one-minute segments where behaviour was observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Contingent behaviour</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Interpersonal behaviour (Child)</td>
<td>2 (3.1%)</td>
</tr>
<tr>
<td>Interpersonal behaviour (Adult)</td>
<td>2 (3.1%)</td>
</tr>
</tbody>
</table>

Each one-minute segment was assessed for the presence of contingent and interpersonal behaviours; these are summarised in Table 3. Each social behaviour could be seen in all three one-minute segments of the majority of clips.

Table 4 reports the intra- and inter-rater reliability estimates for the measurements of duration of gaze and vocalisation by adults and children in a sample of 20 randomly selected one-minute segments. Assessment of gaze by both participants was found to be reliable, with ICC and Spearman correlations in excess of 85% for both intra- (within-) and inter- (between-) rater comparisons.

Duration of vocalisation proved less reliable, though for adult vocalisations, intra- and inter-rater reliability, as assessed by the Spearman Rank correlation was in excess of 80%. The non-Normal distribution of these characteristics may account for the lower ICC estimates of 60–70%. More significant difficulties were observed in assessing the duration of infant vocalisations. In terms of intra-rater reliability, the ICC and rank correlation estimates were high, at 85% and 83%. Between raters, a poor ICC was obtained for infant vocalisation, though the rank correlation showed both raters to order individuals in a consistent manner. This was caused by the two raters having a systematic difference in the scale of measurement; the median (inter-quartile range, IQR) of the estimated duration of vocalisation (in seconds) over the 20 segments were, for rater 1, 2.1 (0.5, 4.7) and, for rater 2, 4.2 (1.7, 8.9). It emerged that the consistent difference between the raters resulted from differing interpretations of when brief episodes of vocalisation close together in time constituted single or multiple utterances. Training of two further observers using a consistent definition produced high inter-rater reliability estimates of 88.8% and 93.9%.

Rater 1 assessed episodes of gaze and vocalisation by both participants in all one-minute segments where this was possible. For each video clip, the consistency of these assessments between segments was evaluated with ICC statistics and mean pairwise Spearman Rank correlation coefficients, as shown in Table 5. Moderate levels of reliability were observed for most assessments, though these were generally better for adult behaviours, at 66% or greater for both gaze and vocalisation. Duration of child vocalisations were found to be least consistent between segments of the same video clip, with a mean rank correlation of only 50%.

Table 4
Intra- and inter-rater reliability estimates.

<table>
<thead>
<tr>
<th>Raters compared</th>
<th>ICC</th>
<th>Rank Corr*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of Gaze (Adult)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-rater</td>
<td>1</td>
<td>97.3%</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>1, 2</td>
<td>89.0%</td>
</tr>
<tr>
<td>Duration of Gaze (Child)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-rater</td>
<td>1</td>
<td>85.8%</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>1, 2</td>
<td>91.2%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Adult)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-rater</td>
<td>1</td>
<td>61.8%</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>1, 2</td>
<td>70.2%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Child)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-Rater</td>
<td>1</td>
<td>84.8%</td>
</tr>
<tr>
<td>Inter-Rater</td>
<td>1, 2</td>
<td>38.3%</td>
</tr>
<tr>
<td></td>
<td>3, 4</td>
<td>88.8%</td>
</tr>
</tbody>
</table>
### Table 5
Between-segment reliability estimates.

<table>
<thead>
<tr>
<th></th>
<th>N segments</th>
<th>ICC</th>
<th>Rank Corr*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of Gaze (Adult)</td>
<td>165</td>
<td>66.2%</td>
<td>66.6%</td>
</tr>
<tr>
<td>Duration of Gaze (Child)</td>
<td>177</td>
<td>62.4%</td>
<td>56.2%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Adult)</td>
<td>129</td>
<td>76.8%</td>
<td>68.6%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Child)</td>
<td>126</td>
<td>58.8%</td>
<td>49.6%</td>
</tr>
</tbody>
</table>

### Table 6
Between-video reliability estimates.

<table>
<thead>
<tr>
<th></th>
<th>ICC All</th>
<th>ICC Meal and Play</th>
<th>ICC Play and Meal</th>
<th>Rank Corr* All</th>
<th>Rank Corr* Meal and Play</th>
<th>Rank Corr* Play and Meal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of Gaze (Child)</td>
<td>56.3%</td>
<td>88.7%</td>
<td>0.0%</td>
<td>13.9%</td>
<td>69.7%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Duration of Gaze (Adult)</td>
<td>27.9%</td>
<td>30.8%</td>
<td>22.1%</td>
<td>13.9%</td>
<td>32.5%</td>
<td>73.3%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Child)</td>
<td>38.5%</td>
<td>12.1%</td>
<td>22.1%</td>
<td>60.1%</td>
<td>44.3%</td>
<td>-16.4%</td>
</tr>
<tr>
<td>Duration of Vocalisation (Adult)</td>
<td>57.1%</td>
<td>16.3%</td>
<td>71.1%</td>
<td>57.4%</td>
<td>43.2%</td>
<td>28.2%</td>
</tr>
</tbody>
</table>

Combining the assessments across segments from each video clip, the consistency between pairs of videos of the same child was similarly evaluated, and is reported in Table 6. In general, there was a much lower degree of reliability when assessing two videos of the same child, particularly when the family was asked to provide either two videos in a play setting, or one video of play and one of a mealtime. However, when both videos were of a meal setting, assessments of gaze were reasonably reliable, particularly for the child, with a 70% rank correlation and 89% estimated ICC. For the adult, the ICC was low, at 31%, though there was a good correlation between videos, at 73%.

Fig. 1 illustrates the total duration of the child’s gaze at the parent’s face and vice versa occurring in the three minutes of video of both play and feeding situations in all video clips where gaze could be accurately assessed for all three one-minute segments. In three minutes of recording, the median, IQR duration of child gaze was 27.2, 14.0–47.0 s for the feed condition, significantly longer than during videos of play (13.7, 4.3–26.9 s) (Mann–Whitney Test, \( p = 0.012 \)). Similarly, adult gaze during mealtimes was significantly longer than during play (94.1, 66.4–104.2 s vs. 43.7, 26.1–62.2 s; \( p = 0.001 \)).

Fig. 2 displays the total duration of child and adult vocalisations in 3 min in all video clips where voices could be accurately assessed throughout.

There were no significant differences between mealtimes and play in either the duration of child (\( p = 0.53 \)) or adult vocalizations (\( p = 0.77 \)).
4. Discussion

Although our overall response rate may appear low, it compares well with community-based studies requesting, for example, DNA samples for epidemiological studies (Watts, 2007). Families were approached “cold” with a request for sensitive personal material for pilot research designed to develop a tool for use in a large study, rather than research with any obvious direct benefit to participants, the National Health Service or society in general. Families had to “opt-in” actively to the study or there would be no further contact from the research team. We believe that response rates within the context of a longitudinal study, where families have a pre-existing trusting relationship with the research team, would be much higher. The high levels of participation in the “Children in Focus” component of the Avon Longitudinal Study of Parents and Children (ALSPAC), in which parents were filmed interacting with their children, provides some support for this contention.1 Some of the small number of families who told us that they did not wish to participate in the study expressed concerns about the security of the video clips and in a larger study it would be desirable to describe data handling procedures to families in more detail. Personal contact with a team member and the opportunity to ask questions could provide reassurance in this respect.

Once cameras were delivered, many families thought it was difficult to find the time to return video clips. We eventually “lost” a substantial proportion of our camcorders and there were substantial delays before clips were returned in many cases. The rate and timing of return of the clips would probably be improved if families were committed to an ongoing study, although this would require confirmation in a pilot cohort study where we would need to liaise closely with families to ensure timely return of data. Some families would have accepted a visit by a researcher to make recordings although this might prove more costly.

Those families who did return clips reported no difficulties in obtaining the video footage and the quality of the returned material was good enough to allow analysis of social behaviours, voice and gaze in the vast majority of clips. The fact that a full view (ie both eyes visible) of parents’ faces was obtained in relatively few segments did not prevent us from assessing approximate direction of gaze, although assessment of visual tracking by parents would be difficult.

A combination of contingent behaviours and interpersonal involvement from both infant and parent was demonstrable in at least one segment of 63/65 (97%) of the video clips. These social behaviours are used in a range of standardised assessments of the parent–child relationship, so such assessment is possible in parent-recorded video material and we have performed structured analyses of parent–infant interactions using standardised scales: the CARE index (Crittenden, 2001) and Mellow Parenting Coding system (Robinson et al., 2009). The results of these analyses will be presented in a separate paper.

Good inter- and intra-rater reliability was achieved for assessment of gaze and parent vocalisation. There was poorer agreement in relation to infant vocalisation, although this can be improved with training. The brief nature of most of the infant vocalisations in the material we obtained seemed to create particular challenges to reliable assessment.

---

1 http://www.alspac.bristol.ac.uk/protocol/Appendix7/Child_Focus2.htm#Selection_of_children.
The reliability estimates (the extent to which observations are consistent) for vocalisation and gaze between one-minute segments within the same video clip were generally fairly high – and predictably higher than reliability between video clips. Consistency was generally higher in meal settings than in the play condition. In epidemiological studies of social communication in infancy, our data support the case for using a partially standardised situation such as mealtime to improve comparability between individuals. It should however be noted that our study was population based, with relatively low levels of abnormal social interaction, and higher levels of agreement might be expected in clinical samples.

5. Conclusions

We have established that it is possible for parents to obtain good quality video material capturing social interaction between them and their babies with cheap video cameras. Accurate analysis of vocal and gaze behaviour as well as other more “interpretable variables” such as interpersonal involvement has proved possible in most of the submitted video material. The technique may be useable in longitudinal epidemiological studies as well as in clinical settings, but this requires confirmation in further pilot work.

Acknowledgements

This study was funded by the Chief Scientist Office of the Scottish Executive Health Department (reference CZG/2/217) and by the NHS Scotland Priorities and Needs Funding Programme. The funding for the camcorders was donated by Philips Electronics, the Swedish Child Neuropsychiatry Science Foundation and the Infant Mental Health Endowment Fund of the Royal Hospital for Sick Children, Yorkhill, Glasgow. The video data were collected by Edwin Burns.

References


