

Correlates of social awareness in the visual prosody of growing children

MARC SWERTS

Tilburg University

Abstract

This article focuses on how growing children use prosody for communicative purposes. Prosody refers to the set of expressive features that do not so much determine what speakers say, but rather how they say it. It includes both auditory features, such as intonation and tempo, and visual features, such as facial expressions. Our central hypothesis is that children, as they grow older, become more socially aware — a process which is reflected in the way they express themselves in prosody. To this end, we present the results of three studies that focus on how children use such features (1) to mark their level of uncertainty, (2) to signal a positive or negative emotion, and (3) to show whether they are being truthful or not. All the studies use a game-based experimental paradigm that is especially suited for analyses of child behaviour. The approach combines controlled elicitations of spontaneous interactions with perception tests that explore how children's expressions are being interpreted. Results of such studies are relevant for pedagogical and diagnostic purposes, and will lead to improvements in child-directed communication systems.

Keywords: prosody, developmental aspects, social awareness, cues to uncertainty, cues to emotion, cues to deception

1. Introduction

When children grow older, they become more proficient in their ability to communicate with other people. As part of this development, they not only increase their knowledge of the words, the pronunciation and grammar of a specific language, but also acquire better skills to use and interpret prosodic structures. Prosody can roughly be defined as the set of features that do not so much determine what people say, but rather how they say it, even though there are examples of how variation in prosody can also change the propositional content of an utterance. Traditionally, the use of the term prosody was restricted to purely auditory

cues, i.e., features that can in principle be measured in the acoustic signal, such as intonation (or speech melody), pauses, rhythm, tempo, loudness and voice characteristics. More recently, the term has been widened to also include a range of visually perceptible forms of body language, such as hand and arm gestures, posture, and facial expressions (Granström and House 2006; Kraemer and Swerts 2009). In a typical conversation, dialogue partners can hear and see each other, so that they naturally pay attention to both auditory and visual features. Consequently, as audiovisual prosody is omnipresent in most interactions, it represents an important communicative device that children need to acquire. In particular, the development of a child's prosodic skills implies that it learns how auditory and visual features can signal linguistically and socially relevant information, e.g. to highlight important words in a discourse, or to signal a speaker's confidence level, or to express different emotions.

Various studies have shown that the acquisition of prosody starts very early in a person's lifetime. For instance, while still in the womb, babies develop a memory for the mother's voice (DeCasper and Fifer 1980), and become sensitive to the rhythmic and intonational patterns of their native language (Mehler et al. 1986; Camras et al. 1992, 1998). They also appear to be genetically predisposed to visual input from a human body, as for instance babies from 1 hour to 3 days can imitate gestures like tongue protrusion and mouth opening after watching an adult producing the same gestures (Meltzoff and Moore 1983). Given that infants are seldomly exposed to events in a single modality at a time, they also learn quickly to unify information coming from different modalities (Spelke 1976). Various people have suggested that biological or physiological factors are very important in understanding the prosodic features of young infants. For instance, even when the respective ambient adult languages differ in the relative proportion of low tones, all infants show a tendency to produce low-ending intonation contours (Snow and Balog 2002), possibly due to simple constraints of the respiratory system (like decreasing air pressure in their lungs throughout an utterance). Also, infants' facial expressions of basic emotions are very similar across the globe, which has sometimes been taken as evidence that such expressions are genetically rooted (Knapp and Hall 2006).

But as a child grows older, the impact of biological or physiological factors on variations in prosody diminishes, as children learn that the features can also be exploited for various communicative purposes. The central working hypothesis of the current article, is that — with age — the prosody of typically developing children becomes more functional and more “other-directed”. In particular, they gradually become more knowledgeable about the conventions that exist within a specific community on how to associate specific expressions with specific pragmatic meanings. This expectation is in line with many theories of cognitive and social development, especially those inspired by Piaget (Goswami 2008), that state that children between ages four and twelve become more socially aware as they grow older. Growing children get exposed to increasingly more varied environ-

ments (family, playground, school), and they find that parents and other people approach them differently when they get older (Swerts and Krahmer 2010). The extent to which growing children change their linguistic and social behaviour has been investigated for a few decades (e.g., Shatz and Gelman 1973; Markman 1973; Goffman et al. 2006), but this research has largely ignored aspects of prosody. Consequently, while it is clear that prosody has added value for human communication, our knowledge of how it develops in growing children is very limited, certainly compared to what we know about other levels of linguistic structure, such as the lexicon and syntax (Diehl and Paul 2009).

Below, we investigate the impact of a social awareness on prosody in three different studies, dealing with expressions of uncertainty (study 1), of emotion (study 2), and of deception (study 3). More detailed accounts of the studies that are described below can be found in Krahmer and Swerts (2005), Shahid et al. (2007, 2008) and Swerts (submitted), respectively. The studies all focus on younger speakers, but differ in the choice of the age groups of the participants being analysed, as we compare 8-year olds with adults in study 1, 8-year and 12-year olds in study 2, and children around the age of 5 in study 3. However, the three studies make use of a similar general approach, which is discussed in the next section.

2. Current set of studies

The three studies discussed below all use a similar 2-step procedure: we first record children in various communicative contexts, and then evaluate their expressions in subsequent perception tests. An important aspect of the first step of this approach, i.e., the elicitation of expressions, is that we use a procedure which guarantees that the recorded materials are representative of the way children express themselves in natural interactions. In that respect, our approach differs from the one used in quite a large number of previous studies. These tended to analyse data with limited ecological validity, being based on portrayed expressions or read-aloud isolated utterances. Characteristics of such data are markedly different from the spontaneous expressions people display in interactions outside the lab. Moreover, production and comprehension tasks that involve portrayal or self-reflection are unsuitable for children, as they require good meta-linguistic and acting skills. Therefore, our paradigm makes use of games, as these have properties highly suitable for studying the audiovisual prosody of children. First, games represent artificial, small universes with their own rules, so that participants can be put in different situational contexts. Second, when people participate in a game, they are interactive, dynamic and engaging; this creates a natural ambiance for spontaneous expressive behaviour (Kaiser and Wehrle 1996; Salen and Zimmerman 2003). And, since games are fun and players tend to enjoy them, there is less risk that they will induce stressful situations. The studies below give examples of games (in the form of a quiz, a card game and a puppet show) that allow us to record children's

expressions in various contexts, which can then be evaluated by independent judges in perception experiments. More specifically, perception experiments allow us to determine to what extent children's expressions can be interpreted by other people, and to what extent they can be meaningfully related to specific communicative functions.

The three studies thus aim to explore different levels of social awareness of children in different age groups. But before we embark on the specifics of these studies, we give some of their general characteristics and limitations. First, we will see that the relation between social development and prosodic variation may take different forms, dependent on the kind of purpose for which prosody is being used. For instance, while we see that with age people become more expressive in their use of prosody to show their confidence level in a question-answering situation (study 1), they tend to become less expressive as they grow older to show basic emotions (study 2). However, these opposing trends in the use of prosody can both be related to a child's increasing knowledge about what is considered to be appropriate behaviour in a specific social context. Second, the three studies described below are primarily based on analyses of Dutch children in order to explore developmental trends in the use of prosody. Based on these studies alone, it is hard to assess the extent to which the reported findings are specific for Dutch children or whether they have more general validity. Still, as we will discuss in the final section of this article, we have replicated study 2 on the display of positive or negative emotions with Pakistani children, which revealed some interesting cultural differences in expressive behaviour, and suggests that developmental patterns in the use of prosody can vary between cultures. And third, all the studies mainly focus on aspects of visual prosody, in particular dealing with cues that are visible in a speaker's face, whereas we pay less attention to auditory features. Facial expressions have been studied quite extensively in the literature, but little is known about the development of such expressions in the growing child. Because of this focus on facial cues, we will not say much on how such visual cues relate to auditory ones, and whether one is more important than the other.

3. Study 1: Cues to uncertainty

3.1. Background

A spoken interaction between two or more dialogue partners is not an exact data transfer process. One cannot always take it for granted that the information contained in a specific utterance is completely accurate. For instance, consider a situation in which one person asks another person a series of questions (e.g., Who won Wimbledon in 2008? What is the capital of Switzerland?). Depending on his or her prior knowledge or interests, the respondent may find such questions easy or difficult to answer. In cases where the question is more complex, this would require a longer memory search which may or may not be successful. As a result, a person

will typically not be able to answer all questions (although sometimes it may feel like the answer lies on the tip of the tongue), and in those cases where an answer is given, it will be associated with a varying degree of uncertainty. Smith and Clark (1993) found that when adult speakers are uncertain about the correctness of their answer, they may indicate this using a variety of prosodic cues including fillers, longer delays, and rising intonation, and using linguistic hedges such as “I guess” or “I think”. In addition, adult speakers also signal their confidence level by means of visual features, such as eyebrow movements, diverted gaze, and marked facial expressions (Swerts and Kraemer 2005).

One could argue that the ability to show cues to uncertainty is related to an increased social awareness. In a question-answering situation, for instance, people may convey this degree of uncertainty to the questioner as a kind of self-presentation; by answering in an uncertain manner a speaker may save face if the answer turns out to be incorrect later on. If a speaker responds to a difficult question with a puzzled look and a hesitant speaking style, that clearly “qualifies” the information he or she tries to convey to an addressee. Conversely, for an addressee, cues to uncertainty shown by a speaker are similarly relevant, as they may give him or her a hint as to whether specific information should be taken with a grain of salt or not. For this reason, it is interesting to compare how children display cues to uncertainty, and how these compare to those of adults.

3.2. *Method*

3.2.1. *Stimuli* Following Smith and Clark (1993), we used Hart’s (1965) method to collect both certain and uncertain speaker utterances from adults and children by asking them a series of factual questions (40 for adults, 30 for children). For adults, the questions came from two sources, where we first selected questions with a one-word answer (e.g., Who wrote Faust? What is the capital of Switzerland?) from a Dutch version of the “Wechsler Adult Intelligence Scale” (WAIS), a standard intelligence test for adults, and added a supplementary list from the Dutch version of the game Trivial Pursuit. For children, the questions were partly taken from a Dutch version of the “Wechsler Intelligence Scale for Children” (WISC). Again, we only selected those questions that allowed for a single word answer, and supplemented these with questions from the Dutch version of Trivial Pursuit for children (e.g., How much is a dozen? Who discovered America?).

Child and adult speakers underwent the same procedure, modulo some small differences detailed below. First, speakers were asked the series of questions by the experimenter, and the speakers’ responses were filmed using a digital camera. The experimenter asked the series of questions one by one, and the pace of the experiment was determined by the participant. Second, after this test, the same sequence of questions was presented again, but now participants only had to indicate how sure they were that they would recognize the correct answer if they would have to

find it in a multiple-choice test. The scores obtained in the second phase are referred to as the Feeling of Knowing (FOK) scores. Adult participants indicated their Feeling of Knowing on a 7-point Likert scale. For second grade children, a standard 7-point Likert scale might cause problems, hence we opted for a 5-point Likert scale using a facial representation of the items with the mouth changing from a sad face (mouth corners pulled down) to a smiling one (mouth corners pulled up). For the purpose of comparison, both Likert scales were recoded to the interval $[0,1]$, with 0 = “I will absolutely not recognize the correct answer in a multiple choice test” and 1 = “I will definitely recognize the correct answer in a multiple choice test”. A speaker’s utterance is said to be uncertain if the corresponding FOK score is low, and certain if the FOK score is high. To stimulate participants to do their best and guess the correct answer in case of uncertainty, they were told that the winner of the game (the person with most answers correct in the first test) would receive a small award (a book token for the adults, and a bag of sweets for the children). In addition, all children received a small reward (a lollipop).

The movie stills in figure 1 show a child and an adult while answering “Who discovered America?” and “Who wrote Faust?”, respectively.

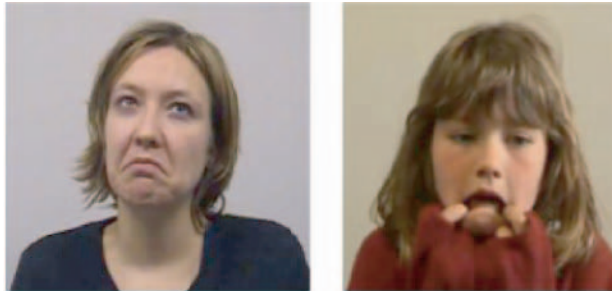


Figure 1 *Movies stills showing uncertain expressions of an adult and a child*

Both the child and adult speaker are unsure of how to respond, but their expressions are quite different. The child uses her fingers to pull down her mouth corners; the adult speaker displays a puzzled look, which is markedly different from her expression when she is more confident. The answers of child and adult sound different as well. The child takes a very long silent pause before she finally responds, whereas the adult, while searching for an answer, produces fillers like “uh” and “uhm”, and additionally delivers her response with a rising question intonation (Swerts and Krahmer 2005).

From the corpus of answers collected in this way, we selected 30 adult utterances and 30 child utterances, each with an equal distribution of high FOK and low FOK scores. The selection was based on the written transcriptions of the responses.

Given the individual differences in the use of the FOK scale, we chose to use the highest score per speaker as an instantiation of high FOK and the two lowest per speaker as representations of low FOK. The original selection of stimuli was random, but utterances were iteratively replaced until the following criteria were met: (1) the original question posed by the experimenter should not re-appear in the speakers response, (2) all the answers should be lexically distinct, and (3) there should be maximally two answers per speaker.

3.2.2. Participants 80 native speakers of Dutch participated as judges, 40 adults and 40 children (20 male and 20 female per group), and all were different from the speakers who participated in the production studies. The children were in two different second grade classes from the St. Jozef School and were 7 or 8 years old. The adults were all students from Tilburg University between 18 and 25 years old.

3.2.3 Perception test The selected clips were presented on a screen where the judges first saw the stimulus ID (1 through 30) and then the actual stimulus. The inter-stimulus interval was 3 seconds for adult judges and 6 seconds for child judges. For all four groups, the experiment was preceded by a short practice session to acquaint judges with the kinds of stimulus materials and the procedure. Judges were instructed to estimate whether speakers were uncertain about their answers or not. Adult judges scored this on a 7-point Likert scale, child judges on a 5-point Likert scale (using the same facial representation as above). For presentation purposes, both scales are recoded to the interval [0, 1], with 0 = “very uncertain” and 1 = “very certain”. These scores are referred to as the Feeling Of Others Knowing (FOAK) scores (Jameson et al. 1993, Brennan and Williams 1995).

3.3 Results and discussion

The overall results are summarized in Table 1. All tests for significance were done using analysis of variance (ANOVA). There was a small but significant main effect of speaker groups ($F_{(1,76)} = 6.574$; $p < .05$). This means that overall child speakers received slightly higher FOAK scores than adult speakers (0.59 and 0.56

Table 1. Average FOAK scores for adult and child judges as a function of speaker type and FOK score

Speaker	FOK	Adult judges	Child judges
Adult	High	0.79	0.66
	Low	0.32	0.47
Child	High	0.70	0.70
	Low	0.44	0.53
Average		0.56	0.59

respectively). Additionally, a main effect of FOK score was found ($F_{(1,76)} = 601.987$; $p < .001$). As one would expect, stimuli with a high FOK score get overall higher FOAK scores than stimuli with a low FOK. A significant two-way interaction was found between speaker and FOK score ($F_{(1,76)} = 26.281$; $p < .01$). Inspection of Table 1 reveals that stimuli from adult speakers receive more disparate FOAK scores than stimuli from child speakers, irrespective of who the judges are. In other words, overall, the FOAK scores are more accurate for adult speakers than for child speakers. There was also a significant interaction between judge and FOK score ($F_{(1,76)} = 62.726$; $p < .01$); differences between FOAK scores for high and low FOK stimuli are larger for adult than for child judges.

The experiment revealed a number of clear differences between adults and children. Overall, we found that FOAK scores assigned to stimuli from adult speakers are more consistent with the original FOK scores than stimuli from child speakers. Adult speakers clearly indicate while answering, whether they are certain about the correctness of their answer or not. In doing so, they can save face when an answer turns out to be incorrect; it is a signal to the questioner saying “this is how certain I think you can be about my answer.” Children in the age group 7-8 appear to be less worried about self-presentation in answering questions. As one piece of anecdotal evidence supporting this view we observe that some of the child participants displayed behaviors that detract from self-presentation, while none of our adult speakers exhibited these behaviors. Specifically, one child repeatedly picked her nose during the experiment, while at least two others pulled down their mouth corners with the fingers (cf. Figure 1). Note, also, that the difference in rating scales used by child and adult raters should not affect the comparison being made here because only high and low values were used.

In a similar vein, various child participants occasionally spent very long times on memory search (more than 30 seconds), without accounting for this delay in any way (and usually still providing an answer that was relevant to the question). The adult participants, by contrast, even though their search space is presumably much larger, hardly ever took more than 10 seconds to answer and in the case of prolonged search almost always indicated this to the questioner via one or more audiovisual cues. In this context it is interesting to observe that delay is a significant cue for child uncertainty while filled pauses are not. A child who does not know an answer immediately tends to engage in a longer memory search, but does not signal this to the listener as an adult would.

4. Study 2: Cues to emotion

4.1. Background

In daily-life interactions, conversants do more than simply exchanging propositional content, but often also tend to show each other how they feel. Prosody is highly relevant for displaying and interpreting emotions. It is generally assumed

that when speakers are in a particular emotional state, this has a clear impact on their auditory and visual cues. Most people find it relatively easy to determine whether a speaker is depressed or elated on the basis of his/her facial expressions and intonation (Darwin 1872; Ekman 1972; Walbott and Scherer 1986; Elfenbein and Ambady 2002; Schmidt and Cohn 2002; Russell et al. 2003; Scherer 2003; Swerts and Hirschberg 2010). While most work has concentrated on either the auditory or visual cues, there has also been some interest in combined vocal-facial expressions, such as in the work by de Gelder and Vroomen (2000), Massaro and Egan (1996), Scherer and Ellgring (2007), Barkhuysen et al. (2010) and Kraemer and Swerts (2011).

The display of emotions is likely to become more other-directed with age. While younger children show their emotions more openly and do not filter them, with age they become less expressive, as they increasingly “internalize” their emotions (Fabes and Martin 1991; Shahid et al. 2007). At the same time, we expect children to become better at producing contextually appropriate expressions, related to their increased social competence (Simonds et al. 2007). While for younger children, emotions may be viewed as the expression of personal experiences, with age, such expressions will increasingly serve as social signals (Fridlund 1994; Manstead 2005). For instance, an older child is generally better at portraying happiness in cases where it actually does not feel that positive, e.g. to produce a polite smile when receiving a present it actually does not like (Simonds et al. 2007). This is consistent with the observation that adults do not always make their auditory or visual expressions congruent with the emotional content of their utterances, e.g., when they deliberately display a happy face with a sad message to signal irony. To see how younger and older children express emotions, we conducted a study in which we compared their audiovisual features after winning or losing a game.

4.2. *Method*

4.2.1. *Stimuli* We developed a simple card game as a tool for inducing emotions in children. When the game starts, players see a row of six cards on the screen of a computer where the number of the first card is visible (‘7’ in the case of the example in Figure 2) and the other five cards are placed upside down so the numbers are hidden. All the numbers on the cards are between 1 to 10 and a number displayed once is not repeated in a particular game. Players have to guess whether the number on the next card will be higher or lower than the previous number. Once players have guessed the number, the relevant card is turned around and the correct answer is visible on the screen. Players are also informed about the correctness or incorrectness of their answer via a characteristic non-speech sound. If players make a wrong guess, the game is finished and they move to the next game. If players guess correctly whether the next number is higher or lower then they are

asked to guess the direction of the next number and players only win a game if they guess the direction of all numbers in the row correctly. If they win a game, they receive a coin. At the end of the experiment, which consists of a number of games, players can exchange the coins they won for a prize.

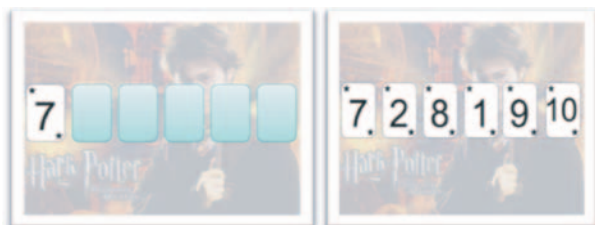


Figure 2. *Beginning and end stage of a card game for emotion induction*

The card games were developed using Microsoft PowerPoint. Appropriate colourful image background and different animations were used to turn cards around in an attempt to make the game attractive for children. During the experiment, children played six games, and could in theory win six coins. However, unknown to the children, each game was completely deterministic, and two different game variants were used. In the first alternative, a rational decision procedure would result in winning the game, and in the second alternative, being rational would result in losing the game. Figure 2 shows an example of the latter: the most probable outcome for the final card would be a “lower” number than 9, but guessing “lower” would make this a losing game. From all of the children who participated in a game playing session, we selected the first response to a winning game (in which they made a correct prediction for the last card) and the first response to a losing game (in which the final guess turned out to be incorrect).



Figure 3. *Movies stills of a child in a winning (left) and losing (right) context*

In total, 48 Dutch children participated in the game (24 8-year olds, 24 12-year olds). Figure 3 gives video stills showing examples of characteristic expressions of a child in a winning or losing context. As can be seen, the child is clearly joyful in the former, and shows signs of disappointment when it notices it has lost the game.

4.2.2. *Participants* 131 adults (university students) participated in the perception experiment, with a roughly equal number of men and women.

4.2.3. *Perception experiment* Stimuli were presented to participants in a random order, in a vision-only format to avoid participants from relying on auditory cues. Participants were invited into a quiet classroom where the computer screen was projected on the classroom wall. They were instructed that they would see children after winning or losing a game and that their task was to determine whether the children had just won or lost their game. Each stimulus was preceded by a number displayed on the screen indicating the upcoming stimulus, and followed by a six second pause during which participants could fill in their score on the answer form.

4.3. Results and discussion

Table 2 summarizes the results. As can be seen, except for the judgments of the older children in a winning context, the percentages correct classifications are generally high, which indicates that participants were able to correctly determine the status of the game based on the display of the emotional response by the children. A repeated measurement analysis of variance shows a significant main effect of Age Group (8 vs. 12 years old) ($F_{(1,127)} = 226.796; p < .001$): inspection of Table 2 indicates that overall the number of correct classifications is higher for the 8 year olds than for the 12 year olds. Another significant main effect was found for Game status (winning vs. loosing) ($F_{(1,127)} = 62.068; p < .001$): overall the percentage of correct classification is higher for losing games. Finally, we also found a significant interaction between Age group and Game Status ($F_{(1,127)} = 127.813; p < .001$). This interaction can be explained by the fact that overall for 8 years old, both winning and losing are mostly recognized correctly, while for 12 years olds, the correct detection of losing games is much higher than that of winning.

Table 2. Mean correct classification results (with standard errors between brackets) for game playing children as a function of age and game status

Age	Game status	Percentage correct
Young	Lost	88.6 (1.7)
	Won	70.1 (2.0)
Old	Lost	75.5 (1.8)
	Won	37.0 (2.0)

These findings are in accordance with the internalization of emotion theory, which says that younger children express their emotions more clearly than older ones. One possible explanation for this is that the emotional expressions of younger children are more direct reflections of a person's internal state, whereas older children appear to be able to "control" these expressions better, and exploit them more as social signals towards others. To explore this further, we have also compared, in a related study, the expressions of children playing the game alone, or in pairs (2 children sitting next to each other). We found overall that the latter situation generally leads to more expressive behaviour in the children, suggesting that such expressions are more than the mere display of an internal mood, as they appear to be affected by the social context (Shahid et al. 2008). However, to explore whether such display rules really differ between the younger and older children, future analyses will have to show whether members from the two age groups also experience the game in a similar way.

5. Study 3: Cues to Deception

5.1. Background

There has been some interest in how prosody may reveal cases of deception. Studies on cues to deception (DePaulo et al. 2003) show that people, in order not to be betrayed as liars, need to be conscious of their auditory and visual expressions. For instance, genuine and fake smiles have been claimed to be minutely distinct in terms of small variations in the eye region and subtle differences in voice quality (DePaulo et al. 2003). However, overall, studies differ in their findings regarding the type and frequency of alleged cues to deception (such as diverted eye gaze). While it generally appears to be very difficult to detect lies on the basis of such expressions, the likelihood of detection may increase dependent on a number of contextual factors and the kind of lie. Our daily social lies are basically impossible to detect, but, somewhat ironically, when there is more at stake, liars are more likely to be detected as such (DePaulo et al. 1996, 2003; Ekman et al 1999).

The analysis of children's ability to deceive and of the way truthful utterances differ from deceptive ones has attracted some scholarly attention. That research is often framed within theory of mind models, under the assumption that the ability to lie requires that one is able to "look inside another person's head", and be able to separate one's own perspective on a given state of affairs from that of another person (Lee et al. 2002; Edelman et al. 2006; Jaswal and Malone 2007; Strömwall et al. 2007). Based on such studies, one could predict that the increasing social competence of growing children allows them to take into account what another person knows or feels, so that they can make their lies consistent with another person's perspective (Talwar and Lee 2002; Talwar et al. 2007). Yet, an opposing

hypothesis would be that this increased social awareness is deficient for a person's ability to deceive; this would be in line with the reported ironic effect that people tend to find it harder to produce a convincing lie when they get more conscious of the fact that they are trying to deceive another person (DePaulo et al. 1996, 2003). To explore such issues, we have looked at truthful and deceptive utterances from younger and older children.

5.2. Method

5.2.1. *Stimuli* In order to obtain truthful and deceptive utterances from children, a specific elicitation procedure was used, based on a computerized version of an animated puppet show (see Figure 4), controlled by a confederate. The show starts with a longer part in which a narrating voice introduces 2 main characters, a prince (the good guy) and a dragon (the bad guy), to the participating child, in a typical fairy-tale plot. The narrator explains to the child that a bad dragon has been terrorizing a far-away country. Luckily, Prince Peter has come up with a plan to capture the dragon, for which he needs the help from the child. The narrator explains that the person who catches the dragon, receives a reward (a bag of gold) from the king. Then the interactive part starts in which child utterances are elicited from exchanges with the 2 main characters of the story, the prince and the dragon. The interactive part contains 2 central scenes designed to elicit minimal pairs of truthful and deceptive utterances from children to be used in a perception test later on. First, the prince appears and tells the child that he wants to capture the dragon, and needs the child's help. He tells the child that he will hide behind a tree (shown on the left of the screen), and that, if the dragon appears, the child needs to tell the dragon that the prince has entered the castle (shown on the right of the screen). Then he hides behind the tree, after which the dragon appears on stage and asks the child where the prince is. After the child has replied with a deceptive "in the castle", the dragon leaves and enters the castle. Then, the prince appears again, tells the child he believes he has heard the dragon, and asks where the dragon is, to which a cooperative child would respond with a truthful "in the castle". Crucially, both the deceptive and truthful scene contains a repeat, in that both the dragon and the prince ask the child to repeat their earlier statement "in the castle". In this way, we obtain 4 versions — first and second attempts of truthful and deceptive utterances — from every participating child of the utterance "in the castle" (or equivalent phrases like "in the tower", or "in the church").

In total 38 children participated in the game (18 boys; 20 girls). The average age of these children was 5 years and 7 months (minimum: 4 years and 10 months; maximum: 6 and 4 months). To explore age differences, the children were categorized into 2 age groups: the younger children (younger than the mean), and older children (older than the mean), which gave about an equal number of younger and older children (52% versus 47%).



Figure 4. Scenes used in the interactive puppet show, taken from the introductory section (top), the truthful interaction with the prince (middle) and the deceptive interaction with the dragon (bottom)

Figure 5 shows some representative stills of a child in a truthful and deceptive context. Note that the child happens to produce a bit of a marked expression (especially visible around the mouth corners) after just having addressed the dragon with a deceptive utterance.



Figure 5. *Movies stills of a child in a truthful (left) and deceptive (right) context*

5.2.2 Participants The participants of the perception study were 60 students from Tilburg University, with one third doing an audiovisual condition, one third an auditory one and one third a visual one (see below).

5.2.3. Perception test The recorded stimuli from the children were presented in pairs to observers in a group experiment, where participants saw the stimuli on a large screen in front of them. The presentation was always set up such that either the truthful or a deceptive utterance by the same child was first shown to a participant on the left of the screen, followed by the other member of a minimal pair shown on the right of the screen. There were 3 between-subject conditions, i.e., one in which the participants could both hear and see the children, one in which they only heard the children (no video), and one in which they only saw moving clips of the children (no sound). In the last case, the participants did not see the actual clips of the children, but saw black screens with the letters A or B representing the first and second members of a pair. The order in which pairs were shown to participants and the order of the members in a pair were fully randomised per group. The task given to the participants was to guess in which member of the pair a child was being deceptive.

5.3. Results and discussion

Table 3 presents the percentages of correct detection of deceptive clips for a number of factors. Given that 50% represents chance level, one can see that the numbers are always close to chance, confirming the observation that the task was a difficult one for the participants. Yet, even when these percentages correct are low, there appear to be significant main effects of condition, age, and attempt. For the

Table 3. *Percentage correct detection of deception (mean, standard error, 95% intervals) and F-statistics for different levels of experimental factors*

Factor	Level	Correct detection	F-stats
Age	Young	52.5 (.10, 50.4–54.7)	$F_{(1,57)} = 11.296$, $p < .01$, $\eta_p^2 = .165$
	Old	57.9 (.13, 55.3–60.5)	
Condition	Auditory	50.8 (.15, 47.8–53.8)	$F_{(2,57)} = 6.916$, $p < .01$, $\eta_p^2 = .195$
	Visual	58.5 (.15, 55.5–61.5)	
	Audiovisual	56.3 (.15, 53.5–59.3)	
Attempt	First	50.9 (.14, 48.1–53.5)	$F_{(1,57)} = 24.934$, $p < .001$, $\eta_p^2 = .304$
	Second	59.5 (.11, 57.2–61.9)	

current article, it is interesting to note that we find an age difference, in that clips from older children are comparatively easier to judge than the ones from the younger ones. Also, with these stimuli, it appears to be the case that the auditory features have weaker cue value than the visual ones, given that observers find it harder to separate truth from deceit when they only have access to auditory cues. And finally, it is interesting to note that the cases where children were prompted to repeat a truthful or deceptive statement were easier to classify than their first attempts. In addition, we found a significant two-way interaction between age and attempt ($F_{(1,57)} = 14.961$, $p < .001$, $\eta_p^2 = .208$), which can be explained by looking at table 4. The table shows that second attempts get better detection results for both younger and older children, but the difference between first and second attempts is bigger for the older ones.

The age effect may seem surprising at first sight, as one might expect older children to be better in deceiving as they are better in judging another person's perspective. But maybe that increased social awareness is actually causing an ironic "pink elephant" effect (Wardlow Lane et al. 1996): the fact that they are aware that they are being observed, makes them realise more that they are trying to deceive. It has been reported before that more conscious lies are easier to detect as such than the innocent daily social lies. Along the same lines, that might explain why second attempts to lie become easier to detect than the children's first attempts, given that the second attempts make children more aware of what they are trying to do. Obviously, more work is needed here as well, given that an alternative hypothesis could be that older children show more cues to deception, because they view the cartoons as artificial characters who cannot see them, so that they have less reason to "hide" their lies.

A possible confound of the current study is that the deceptive utterances were produced towards the bad character (the dragon), and the truthful ones to the good character (the prince). Theoretically, this could mean that the cues used in the perception experiment were not so much correlates of telling a lie, but rather signs that the children were a bit scared or nervous because of the dragon, which was then later interpreted as cues to deception. This in itself is not very unusual, of

Table 4. Percentage correct detection of deception (mean, standard error, 95% intervals) as a function of the factors age and attempt

Age	Attempt	Mean	Standard error	95% intervals
Young	First	50.4	.16	47.2–53.3
	Second	54.6	.14	51.9–57.4
Old	First	51.3	.17	47.9–54.7
	Second	64.4	.16	61.2–67.7

course, as lies are usually accompanied by other negative feelings, such as uncertainty, regret or fear. Nevertheless, one could solve this by switching roles and let children lie or tell the truth to equally neutral figures, or even invent a story where they have to deceive the good character and be truthful to the bad one (e.g., to protect them in one way or another).

6. General discussion and conclusion

This article has reported on the development of audiovisual prosody in children. The central hypothesis addressed in different studies was that this prosody shows correlates of a child's social awareness, which is likely to become more developed with age. We indeed found from analyses of expressions of uncertainty (study 1), of emotion (study 2) and of deception (study 3) that social awareness may explain developmental patterns in audiovisual behaviour. Children do not show their confidence level as clearly as adults do, presumably because they still need to learn that showing one's uncertainty is a handy face-saving strategy, and is communicatively relevant for an addressee. In addition, older children are less expressive than younger ones about positive or negative emotions, maybe because such emotional cues are more internalised in the older children, and used more as signals that can to some extent be manipulated for social purposes. And finally, older children have a harder time hiding their lies, which could be due to the fact that they are more socially aware of the addressee whom they try to deceive, which makes them more conscious about their lies.

There are a number of potentially interesting follow-up studies. First, only our study on cues to uncertainty has investigated perceptual skills of children, but it could be worthwhile to extend this to the other studies as well. That is, to be regarded as fully competent in audiovisual prosody, one should not only be able to produce appropriate expressions that fit specific linguistic and social contexts, but also be able to correctly interpret audiovisual expressions of another person's messages (Krahmer and Swerts 2009). Whereas adults' productive skills in audiovisual prosody are generally on a par with their receptive skills, it is not clear how these two relate in a growing child. Developmental studies of prosody sometimes reveal results that deviate from what we know about other kinds of development.

That is, for many aspects of cognitive development, it has repeatedly been shown that comprehension precedes production, e.g., when an infant is able to produce 10 words, the number of words comprehended ranges from 30 to 182 (Benedict 1979). The reason for this inequality between production and reception is that input logically precedes output, as infants learn their linguistic and social skills largely from imitating their primary caretakers. However, in line with what has been reported for the development of discourse phenomena, such as the use of pronouns (Spenader et al. 2009), there is a debate as to whether this advantage in comprehension skills also holds for the acquisition of audiovisual prosody. Indeed, some older studies have shown that children's prosodic productions are more advanced than their prosodic understanding (Cutler and Swinney 1987), in the sense that children are already able to produce appropriate patterns, but still fail to interpret them in others. It is not known whether and how this asymmetry between production and comprehension holds for all aspects of audiovisual prosody (Wells et al. 2004).

In addition, it could also be worthwhile to look into possible cultural differences in the development of audiovisual prosody, given that the conventions to show or hide expressions may vary across the globe. We have replicated our card game (study 2) with 8- and 12-year old Pakistani, and found some intriguing differences with the Dutch data. We elicited winning and losing situations following exactly the same procedure as sketched above, and presented the recorded clips to observers in subsequent perception tests. We found that overall Pakistani children are more expressive than the Dutch ones, so that their clips are easier to classify (both by Dutch and Pakistani observers). But we found an additional cultural effect as well. As we saw above, for Dutch children, it appears that losing is easier to detect than winning, which is true for both 8 and 12 year olds. This is completely different for the Pakistani children, where in all variants the classifications were correct in most of the cases, and where winning the game is somewhat easier to detect than losing, both for 8 and for 12 year old children. Interestingly, this pattern (Dutch signal negative emotions more clearly, Pakistani signal positive emotions more clearly) was also found in an analysis of how adult speakers signal emotions in a completely different setting (Shahid et al. 2007). Further work needs to be done to explore why it is that differences exist between the Dutch and Pakistani participants, and whether indeed they might relate to differences in social and/or cultural norms. Possibly, the differences may also relate to the different role within a specific linguistic structure at large. For instance, while in languages like Dutch and English, certainty, truthfulness and other 'evidential' values are largely encoded by prosodic variation, prosody may be less salient in languages that have more morphosyntactic means of marking such features. See Swerts et al. (2002) and Swerts (2007) for a more elaborate discussion of such issues.

Results from these studies have pedagogical and diagnostic implications; they can become part of methods to diagnose and help children who have comparatively poor communicative skills, and to teach children to make themselves better

understood in a linguistically and socially acceptable way (Lloyd et al. 1998). This is especially relevant for children with autism. Autism is currently believed to affect more than 0.6% of people, and is three times as common in males (Hill and Frith 2003). While children with autism cannot be cured of autism, they can be taught some behavioural skills by explicit training procedures. For instance, there is recent work which shows that children with autism display more social behaviour after repeated imitation sessions (Field et al. 2001). Unfortunately, a recent overview paper by Peppé (2009) shows how little attention prosody has so far received in studies of speech-language pathology, despite the intuition that deficiencies in prosodic competences are likely to have negative repercussion for a person's ability to communicate with others (see also Swerts 2009). People who do not master the prosodic rules of a language, experience problems expressing themselves in an acceptable way, and find it difficult to interpret the expressions of another person. Our games-based paradigm is especially suitable, as it naturally elicits different linguistic and social contexts. By playing games, people with autism can be trained to recognise different contexts, to which they can adapt their audiovisual expressions.

Acknowledgments

The research presented in this article was conducted as part of the VIDI-project "Functions of Audiovisual Prosody (FOAP)", sponsored by the Netherlands Organisation for Scientific Research (NWO), see foap.uvt.nl. Many thanks to Emiel Kraemer, Suleman Shahid, Lennard van de Laar, Judith Schrier, Jorien Scholze, Kim Smulders, Nicole Hobbelen, Sjoukje Houbers and Manon van Dijk, for various kinds of collaborations which have led to the three studies of this article.

Correspondence e-mail address: m.g.j.swerts@uvt.nl

References

- Barkhuysen, Pashiera, Emiel Kraemer & Marc Swerts. 2010. Cross-modal and incremental perception of audiovisual cues to emotional speech. *Language and Speech* 53(1). 3–30.
- Benedict, Helen. 1979. Early lexical development: Comprehension and production. *Journal of Child Language* 6. 183–200.
- Brennan, Susan E. & Maurice Williams. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive state of speakers. *Journal of Memory and Language* 34. 383–398.
- Camras, Linda A., Harriet Oster, Joseph Campos, Kazuo Miyake & Donna Bradshaw. 1992. Japanese and American infants' responses to arm restraint. *Developmental Psychology* 28. 578–583.
- Camras, Linda A., Harriet Oster, Joseph Campos, Rosemary Campos, Tatsuo Ujiie, Kazuo Miyake, Lei Wang & Zhaolan Meng. 1998. Production of emotional facial expressions in European American, Japanese, and Chinese infants. *Developmental Psychology* 34 (4). 616–628.

- Cutler, Anne & David A. Swinney. 1987. Prosody and the development of comprehension. *Journal of Child Language* 14. 145–167.
- Darwin, Charles. 1872 [1998]. *The expression of the emotions in man and animals*, 3rd edn. Oxford: Oxford University Press.
- DeCasper, Anthony J. & William P. Fifer. 1980. On human bonding: Newborns prefer their mother's voices. *Science* 208. 1174–1176.
- DePaulo, Bella M., Deborah A. Kashy, Susan E. Kirkendol, Melissa M. Wyer & Jennifer A. Epstein. 1996. Lying in everyday life. *Journal of Personality and Social Psychology* 70(5). 979–995.
- DePaulo, Bella M., James L. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton & Harris Cooper. 2003. Cues to deception. *Psychological Bulletin* 129(1). 74–118.
- Diehl, Joshua J. & Rhea Paul. 2009. The assessment and treatment of prosodic disorders and neurological theories of prosody. *International Journal of Speech-Language Pathology* 11(4). 287–292.
- Edelstein, Robin S., Tanya L. Luten, Pauk Ekman & Gail S. Goodman. 2006. Detecting lies in children and adult. *Law and Human Behavior* 30. 1–10.
- Ekman, Paul. 1972. *Emotion in the human face*. Oxford: Pergamon Press.
- Ekman, Paul, Maureen O'Sullivan & Mark G. Frank. 1999. A few can catch a liar. *Psychological Science* 10(3). 263–266.
- Elfenbein, Hillary A. & Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128. 203–235.
- Fabes, Richard A. & Carol L. Martin. 1991. Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin* 17. 532–540.
- Field, Tiffany, Tory Field, Chris Sanders & Jacqueline Nadel. 2001. Children with autism display more social behaviors after repeated imitation sessions. *Autism: The International Journal of Research and Practice* 5(3). 317–323
- Fridlund, Alan J. 1994. *Human facial expression: An evolutionary view*. New York: Academic Press.
- de Gelder, Beatrice & Jean Vroomen. 2000. The perception of emotions by ear and by eye. *Cognition and Emotion* 14. 289–311.
- Goffman, Lisa, Lori Heisler & Rahul Chakraborty. 2006. Mapping of prosodic structure onto words and phrases in children's and adults' speech production. *Language and Cognitive Processes* 21. 25–47.
- Goswami, Usha. 2008. *Cognitive development: The learning brain*. Hove, NY: Psychology Press.
- Granström, Björn & David House. 2006. Measuring and modeling audiovisual prosody for animated agents. *Speech Prosody 2006*, Dresden.
- Hart, Joseph T. 1965. Memory and the feeling-of-knowing experience. *Journal of Educational Psychology* 56. 208–216.
- Hill, Elisabeth L. & Uta Frith. 2003. Understanding autism: Insights from mind and brain. *Philosophical Transactions of the Royal Society of London B* 358(1430). 281–289.
- Jameson, Anthony, Thomas O. Nelson, R. Jacob Leonesio & Louis Narens. 1993. The feeling of another person's knowing. *Journal of Memory and Language* 32. 320–335.
- Jaswal, Vikram K. & Lauren S. Malone. 2007. Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development* 8(3). 263–283.
- Kaiser, Susanne & Thomas J. Wehrle. 1996. Situated emotional problem solving in interactive computer games. In N. H. Frijda (ed), *9th Conference of the International Society for Research on Emotions*, Toronto. 276–280.
- Knapp, Mark L. & Judith A. Hall. 2006. *Nonverbal communication in human interaction*. Thomson: Wadsworth.
- Krahmer, Emiel & Marc Swerts. 2005. How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech* 48. 29–53.
- Krahmer, Emiel & Marc Swerts (eds.). 2009. Audiovisual prosody. [Special issue]. *Language and Speech* 52(2-3).
- Krahmer, Emiel & Marc Swerts. 2011. Audiovisual expressions of emotions in communication. In Joyce Westerink, Martijn Krans & Martin Ouwerkerk (eds.), *Sensing emotions in context: The im-*

- pect of context on behavioral and physiological experience measurements* (Philips Research Book Series, Vol. 12). Berlin: Springer Verlag.
- Lee, Kang, Catherine Ann Cameron, Joanne Doucette & Varun Talwar. 2002. Phantoms and fabrications: Young children's detection of implausible lies. *Child Development* 73(6). 1688–1702.
- Lloyd, Peter, Sandi Mann & Ian S. Peers. 1998. The growth of speaker and listener skills from five to eleven years. *First Language* 18. 81–103.
- Manstead, Antony S. R. 2005. The social dimensions of emotions. *The Psychologist* 18(8). 484–487.
- Markman, Ellen M. 1973. Factors affecting the young child's ability to monitor his memory. Philadelphia: University of Pennsylvania Ph.D thesis.
- Massaro, Dominic W. & Peter B. Egan. 1996. Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review* 3. 215–221.
- Mehler, Jacques, Ghislaine Lambertz, Peter W. Juszyk & Claudine C. Amiel Tison. 1986. Discrimination de la langue maternelle par le nouveau-né. *Comptes rendus de l'Académie des Sciences de Paris* 303, série III (15). 637–640.
- Meltzoff, Andrew N. & M. Keith Moore. 1983. Newborn infants imitate adult facial gestures. *Child Development* 54. 702–709.
- Peppé, Sue. 2009. Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology* 11(4). 258–271.
- Russell, James A. G., Jo Anne Bachorowski & José Miguel Fernández-Dols. 2003. Facial and vocal expressions of emotion. *Annual Review of Psychology* 54. 329–349.
- Salen, Katie & Eric Zimmerman. 2003. *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.
- Scherer, Klaus G. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40. 227–256.
- Scherer, Klaus G. & Heiner Ellgring. 2007. Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* 7. 158–171.
- Schmidt, Karen & Jeffrey F. Cohn. 2002. Human facial expressions as adaptations: Evolutionary questions in facial expressions. *Yearbook of Physical Anthropology* 44. 3–24.
- Shahid, Suleman, Emiel Kraemer & Marc Swerts. 2007. Audiovisual emotional speech of game playing children: Effects of age and culture. *Interspeech 2007*, Antwerp, Belgium. 2681–2684
- Shahid, Suleman, Emiel Kraemer & Marc Swerts. 2008. Real vs. acted emotional speech: Comparing South-Asian and Caucasian speakers and observers. *4th International Conference on Speech Prosody*, Campinas, Brazil. 669–772.
- Shatz, Marilyn W. & Rochel Gelman. 1973. The development of communication skills: Modifications in the speech of young children as a function of the listener. *Monographs of the Society for Research in Child Development* 38 (5).
- Simonds, Jennifer, Jessica E. Kieras, M. Rosario Rueda & Mary K. Rothbart. 2007. Effortful control, executive attention, and emotional regulation in 7–10 year-old children. *Cognitive Development* 22. 474–488.
- Smith, Vicki L. & Herbert H. Clark. 1993. On the course of answering questions. *Journal of Memory and Language* 32. 25–38.
- Snow, David P. & Heather L. Balog. 2002. Do children produce the melody before the words? A review of developmental intonation research. *Lingua* 112. 1025–1058.
- Spelke, Elizabeth S. 1976. Infants' intermodal perception of events. *Cognitive Psychology* 8. 553–560.
- Spenader, Jennifer, Erik Jan Smits & Petra Hendriks. 2009. Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language* 36(1). 23–52.
- Strömwall, Leif A., Pär Anders Granhag & Sara Landström. 2007. Children's prepared and unprepared lies: Can adults see through their strategies? *Applied Cognitive Psychology* 21. 457–471
- Swerts, Marc. 2007. Contrast and accent in Dutch and Romanian. *Journal of Phonetics* 35(3). 380–397.
- Swerts, Marc. 2009. The relevance of visual prosody for studies in language and speech-language pathology. *International Journal of Speech-Language Pathology* 11(4). 282–286.

- Swerts, Marc. Submitted. About pink elephants and green dragons: Correlates of children's increased awareness of being deceptive.
- Swerts, Marc & Julia Hirschberg. 2010. Prosodic predictors of upcoming positive or negative content in spoken messages. *Journal of the Acoustical Society of America* 128(3). 1337–1344.
- Swerts, Marc & Emiel Krahmer. 2005. Audiovisual prosody and feeling of knowing. *Journal of Memory and Language* 53. 81–94.
- Swerts, Marc & Emiel Krahmer. 2010. Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics* 38. 197–206.
- Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654.
- Talwar, Varun & Kang Lee. 2002. Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development* 26(5). 436–444.
- Talwar, Varun, Heidi M. Gordon & Kang Lee. 2007. Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology* 43(3). 804–810.
- Walbott, Harald G. & Klaus G. Scherer. 1986. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology* 51. 690–699.
- Wardlow Lane, Liane, Michelle Groisman & Victor S. Ferreira. 2006. Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science* 17(4). 273–277.
- Wells, Bill, Sue Peppé & Nata K. Goulandris. 2004. Intonation development from five to thirteen. *Journal of Child Language* 31. 749–778.