Computers in Human Behavior 29 (2013) 654–664

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Using non-verbal cues to (automatically) assess children's performance difficulties with arithmetic problems

Marije van Amelsvoort *, Bart Joosten, Emiel Krahmer, Eric Postma

Tilburg University, Department of Humanities, Tilburg Center for Cognition and Communication (TiCC), P.O. Box 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Available online 16 January 2013

Keywords:
Facial expressions
Arithmetic problems
Performance difficulty
Affective tutoring systems

ABSTRACT

Intelligent tutoring systems often make use of students' answers to adapt instruction or feedback on a task. In this paper, we explore the alternative possibility of adapting a system based on the perceived affective and cognitive state of a student. A system can potentially better adapt to the needs of each individual student by using non-verbal behavior. We used a new experimental paradigm inspired by 'brain training' software to collect primary school children's answers to easy and difficult arithmetic problems and made audiovisual recordings of their answers. Adult observers rated these films on perceived difficulty level. Results showed that adults were able to correctly interpret children's perceived level of difficulty, especially if they saw their face (compared to hearing their voice). They paid attention to features such as 'looking away', and 'frowning'. Then we checked whether we could also *automatically* predict if the posed problem was either easy or difficult based on the first second of their response. This 'thin-slice analysis' could correctly predict the difficulty level in 71% of all cases. When trained on sufficiently many recordings, Adaptive Tutoring Systems should be able to detect children's state and adapt the difficulty level of their learning materials accordingly.

1. Introduction

Imagine yourself in primary school. The teacher has just explained a new type of arithmetic problem and you had trouble understanding it. Now you start working on the arithmetic problems yourself with the help of a computer program. You cannot solve the first problem and the computer gives you a somewhat easier problem. It does not make you less frustrated though; you feel like you'll never be able to do this! Luckily, the teacher sees your face and comes up to your table to ask you how you are doing. You feel a bit better already.

Computers have become an intrinsic part of education. Many software tools have been developed to train and assess children's knowledge and skills. One of the advantages of using a computer to provide children with tasks is that tasks can be adapted to the individual child. Adaptive systems can provide learning materials that fit the needs of learners. Almost all adaptive systems adapt to a learner's *knowledge* (Brusilovsky, 1996). To be able to do this, a system has to recognize a learner's knowledge state and update the model of this particular learner accordingly. Subsequently, it adapts the content of the system to the learner's current knowledge state. Programs have been developed that adapt the difficulty level when children give many wrong answers. However, whether the answer is right or wrong is not the only thing important for learning. For example, a learner can be engaged in a task and enjoy trying things even when making mistakes, because the task is still in his or her zone of proximal development (Vygotsky, 1978). On the other hand, it may be good to adjust the difficulty level when the learner is showing signs of boredom or increased frustration (Kapoor, Mota, & Picard, 2001) when making errors.

For a computer to be able to adjust to learners' needs, it needs to 'understand' them, which is more than adapting to answers on questions. In human communication, Quintilianus already argued in the first century AD that *how* things are said is at least equally important as *what* is said. The face has been called the primary source of information for someone's affective state (Knapp & Hall, 2006), and in interaction, facial expressions support the information a speaker wants to convey (e.g., Barkhuysen, Krahmer, & Swerts, 2005; Ekman, 1979). In a voice too, we may detect boredom, frustration, uncertainty, and other affective or cognitive states (e.g., Brennan & Williams, 1995; Dijkstra, Krahmer, & Swerts, 2006).

Children's non-verbal behavior may show whether a task is too easy or too difficult for them, and enable a computer to adapt the level of task-difficulty accordingly. Reading non-verbal behavior may be especially important in young learners, since these children may not yet be able to express themselves verbally very well. This makes their non-verbal reactions a valuable source of information. It seems easier for children to express themselves non-verbally,

* Corresponding author. Tel.: +31 13 466 3579; fax: +31 13 466 2892.
E-mail address: m.a.a.vanamelsvoort@uvt.nl (M. van Amelsvoort).

0747-5632/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2012.10.016

also in school tasks. For example, Alibali (1999) found that when children generate new problem-solving strategies, they often first show gestures expressing these strategies before being able to verbalize them. Another reason to use non-verbal information in school tasks is that children's facial expressions may reveal their level of meta-cognitive awareness. Meta-cognition is important for regulating one's own learning. Smith and Clark (1993) for example showed that people signal uncertainty in factual question–answering situations by a variety of verbal prosodic cues, and Swerts and Krahmer (2005) extended this finding to the visual domain.

Expert teachers seem to be able to infer children's affective state and modify their pedagogical tactics accordingly (Goleman, 1995). However, until now it is not exactly known on what grounds teachers evaluate how children are doing while performing school tasks. They probably use non-verbal expressions, but the exact nature of the information used and how they give rise to a course of action are still open questions (D'Mello et al., 2005). This lack of understanding is partly due to the fact that drawing inferences from non-verbal behavior in general is largely an unconscious process.

The aim of our study is to elucidate the nature and recognition of non-verbal cues to perceived task difficulty in children. In other words, *do non-verbal cues give information about perceived difficulty of school tasks in children?* We approached this question from three angles: first, from a subjective angle; by investigating whether people can rate perceived difficulty based on non-verbal cues (and if so, how); second, from an objective angle, by manually coding features of non-verbal behavior; and third, from an automatic angle, by researching whether a computer can create a model for predicting perceived difficulty level. Subsidiary questions are whether non-verbal cues are best shown in the face or in the voice, and whether there are age differences in expressions of perceived difficulty. We chose to look at arithmetic problems to answer these questions, because arithmetic problems are short (so it is relatively easy to collect sufficient data), have straightforward answers, and provide standardized ways to decide what difficulty level a child should have achieved in arithmetic at a certain age.

1.1. Non-verbal cues and learning

Facial expressions can be defined as gestures executed with the facial muscles. They are often associated with an individual's affective state, showing feelings and emotions (e.g., joy, surprise, frustration), but they may also reflect the cognitive state (e.g., indicating that someone is thinking) or have a conversational function (e.g., to indicate a question).

When studying facial expressions of emotion, many researchers focus on so-called "basic emotions", such as joy, fear, anger, disgust and sadness. However, emotions that are associated with learning may be different from basic emotions. Adolphs (2002) distinguishes between several types of emotions; besides *basic ones*, he distinguishes *social emotions*, such as pride and embarrassment, *moods or background emotions*, such as cheerfulness and anxiety, and *motivational states*, such as reward and punishment. These emotions appear to be much closer related to learning than basic emotions. Previous research has shown that affective states such as frustration, boredom, interest and confusion can occur in learning (Craig, Graesser, Sullens, & Gholson, 2004). D'Mello et al. (2005) argue that a learner experiences a variety of emotions in learning, e.g. surprise or curiosity in case of extreme novelty of learning material, or anger or frustration when important goals are blocked by this novelty. Positive emotions such as satisfaction or joy can occur when learning goes well.

However, to detect whether a child finds an arithmetic problem easy of difficult we may not need to recognize specific, detailed emotions. It may well be that particular non-verbal cues are indicative of performance problems.

Besides facial expressions, our voice can also tell people how we feel. This is particularly true for our prosody, which can be described as 'a whole gamut of features that do not determine what people are saying, but rather *how* they are saying it' (Ladd, 1996). Although some researchers have extended prosody to include visual features (e.g., Swerts & Krahmer, 2005), prosody is often limited to acoustic characteristics such as pitch, voice quality, loudness, rhythm, and pauses (e.g. 't Hart, Collier & Cohen, 1990). It has been shown that this kind of information can be helpful for detecting affective and cognitive states (e.g., Scherer, 2003). The combination of facial expressions and voice characteristics may improve detection of affective state in learners. Among many others, De Silva, Miyasato, and Nakatsu (1997), for instance, combined auditory and visual information for emotion recognition. Yoshitomi, Kim, Kawano, and Kitazoe (2000), in trying to build a robot that reacts to emotional states from humans, found that both robots and human beings recognize basic emotions better if auditory prosody and facial expressions are combined. Huang, Chen, Tao, Miyasato, and Nakatsu (1998) also found that by using both modalities, it was possible to achieve higher recognition rates than either modality alone. However, much less is known about detection of social or interactive emotions, which may be more subtle and difficult to detect (Back, Jordan, & Thomas, 2009).

1.2. Age differences in showing facial expressions

Using non-verbal behavior to get information about learners' affective state may be specifically relevant when learners are children. Children are in general more expressive than adults, and Thompson (1994) argues that younger children are more expressive than older children. There is some debate about the reason for this. On the one hand, young children may be more expressive in their face as a compensation for their limited verbal abilities (Doherty-Sneddon & Kent, 1996). Thus, when children grow older they might not need to express themselves non-verbally as much as before because of their improved verbal skills. On the other hand, they might also have adapted their facial expressions to support the verbal information they want to convey. In addition, children may learn how to control their facial expressions for social reasons while growing up (Krahmer & Swerts, 2005).

1.3. Software for automatic detection of affective state

Novice teachers could greatly benefit from the knowledge of how to detect children's inner state, but knowledge on facial and voice expressions in learning could also be used to improve educational software. Many software tools have been developed to train and assess children's knowledge and skills. One of the advantages of using a computer to provide children with tasks is that tasks can be adapted to the individual child. Adaptive systems can provide learning materials that fit the needs of the learners.

Researchers are starting to investigate the detection of learner's affective states and other learner characteristics in order to incorporate these aspects into educational software (Craig, D'Mello, Witherspoon & Graesser, 2007; Kapoor et al., 2001; Sarrafzadeh, Alexander, Dadgostar, Fan, & Bigdeli, 2008). For example, Graf, Liu, Kinshuk, Chen, and Yang (2009) showed connections between students' learning styles and cognitive traits, and advocate the use of these connections in web-based educational systems to tailor to the needs of students. These additional characteristics make the system more adaptable to individual needs and current affective states. Systems that take emotions into account are called Affective Tutoring Systems (ATSs). However, automatic detection of affective states is still maturing. Sarrafzadeh et al. report on the development of an ATS for mathematics, called "Easy with Eve", but to our knowledge this ATS has not yet been implemented in real

school situations. These kinds of systems need lots of input to 'teach' them to recognize affective states (Littlewort et al., 2011).

Our study aims to contribute to this line of research, paving the way for automatic detection of non-verbal expressions in learning by collecting suitable training data, and by developing an automatic method to detect performance difficulties based on these data.

1.4. Research questions and organization of the paper

The main research question addressed in this paper is: Do non-verbal cues give information about perceived difficulty of arithmetic problems in children? We investigate this question with several sub questions:

1. Can adults interpret non-verbal cues to infer whether children are dealing with an easy or difficult problem?
2. What non-verbal cues can be detected in children's faces when they perform easy and difficult arithmetic problems?
3. Is a combination of both facial expressions and voice characteristics better for detection of performance problems in learning than either modality alone?
4. Are older children less expressive in their face while solving arithmetic problems than younger children?
5. To what extent can non-verbal cues be automatically detected?

The research questions are answered by discussing five studies. In Section 2, we explain the production study in which children of different age groups answered easy and difficult arithmetic problems using a new experimental paradigm, while being recorded on film. In Section 3, we discuss two perception studies in which adults were shown selected fragments in three different modes (video only, audio only, video and audio) and were asked to indicate the difficulty level the child perceived. This was done with both the entire response (Section 3) and with only the answer (Section 4). In Section 5, we discuss what types of expressions were shown in the faces of the children. In Section 6, we investigated whether automatic detection of facial expressions is possible, based on the data we collected. In our general discussion in Section 7, we combine these different studies and discuss implications for education and further research.

2. Data-collection

To collect facial expression data in learning, we elicited responses to easy and difficult arithmetic problems from children of two age groups. Arithmetic problems were chosen because they have straightforward answers and there are clear guidelines on the arithmetic level a child at a certain age should be on.

2.1. Method

2.1.1. Design

We employed a mixed 2 × 2 design with grade (second grade, fifth grade; in Dutch group 4 and group 7) as between-subjects variable, and level of difficulty (easy, difficult) as within-subjects variable. The order of the arithmetic problems was randomly varied to prevent order-effects.

2.1.2. Participants

Fifty-eight children from a primary school in the Netherlands participated in this study; 29 from second grade (which is group 4 in the Dutch school system) and 29 from fifth grade (group 7; we will use the Dutch terms in this article). The 14 boys and 15 girls in group 4 had a mean age of 7 years and 6 months (*SD* = 6 months), and the 14 boys and 15 girls in group 7 had a

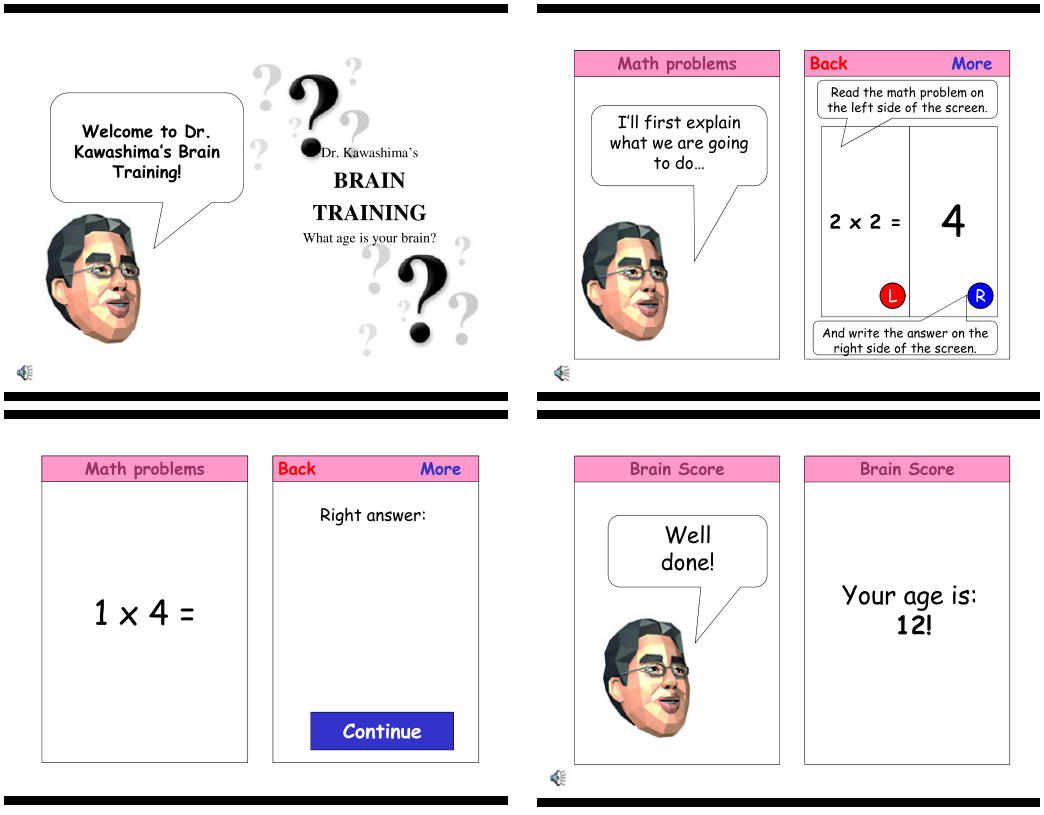

Fig. 1. PowerPoint slides showing welcome, explanation, a simple arithmetic problem and age calculation. Text is translated from Dutch into English for this article.

Fig. 2. Facial expressions in answering arithmetic problems (clockwise from top left: funny face, laughing, averted gaze, frowning).

mean age of 10 years and 8 month (*SD* = 6 months). Children's parents were informed about the experiment beforehand and returned a consent form for their child's participation and usage of recorded material for research purposes.

2.1.3. Material

A PowerPoint presentation (Fig. 1) was developed with arithmetic problems at an easy and a difficult level. The PowerPoint resembled Dr. Kawashima's Brain Training, a popular educational game developed by Nintendo DS. Many children in the Netherlands are familiar with this game and played it regularly themselves at the time of the data collection. In this game, as in ours, people have to solve arithmetic problems as fast as they can. Children's reactions to the arithmetic problems were recorded with a videocamera placed behind the laptop with the PowerPoint. The problems were taken from the Tempo Test Rekenen [Arithmetic Speed Test], an official test that children take regularly in the Dutch school system (De Vos, 1994). The difficulty level was based on norms of what the children's level should be. Thus, half of the problems for each grade were taken from a level below the children's current ability level, and half of the problems were taken from a level above their current ability level.

2.1.4. Procedure

The children were informed that the researchers were evaluating a new version of Dr. Kawashima's Brain Training, and were asked to perform a task, consisting of arithmetic problems. They carried out this task one by one, in a separate room in the school. The experimenter first talked to the children to make them at ease, and told them that the task was a game and did not involve getting a grade. Then children watched the PowerPoint in which Dr. Kawashima first explained the game and then gave 12 easy and 12 difficult arithmetic problems. The problems were blocked and the order of the blocks was counterbalanced, such that half of the children started with the easy and half with the difficult problems. At the end of the slides, Dr. Kawashima "calculated" the children's intellectual age, and systematically gave them a higher age than their real age. This was done to prevent the children from feeling insecure due to the 12 problems far above their ability level. At the end of the experiment, children were asked to indicate the general level of difficulty and fun they experienced in playing the game. This was done with a five-point scale consisting of facial representations, with the items changing from a sad face (mouth corners pulled down) to a smiling one (mouth corners pulled up). These scales are often used with children (e.g., Lockl & Schneider, 2002; Read, MacFarlane, & Casey, 2002). All children received a small treat to thank them for their participation.

2.2. Results

Fig. 2 shows representative stills of children's reactions after receiving an easy or a difficult arithmetic problem. Overall, the game worked quite well. All children indicated that they liked the task rating it on average with $M = 4.17$ ($SD = 0.60$), on a five-point scale ranging from 'I did not like the game at all' to 'I liked the game very much'. The majority of children (34 out of 58) rated the task 'not easy/not difficult', $M = 2.88$, $SD = 0.68$, on the five-point 'smiley' scale, which suggests that the arithmetic problems taken from the Arithmetic Speed Test were indeed both easy and difficult for their level. There was no significant age difference in the amount of fun or level of difficulty children experienced.

Furthermore, the data gathered seem rich in facial expressions. Informal observations reveal differences in facial expressions between the easy and difficult problems and between age groups, which we attempt to validate in a perception experiment.

3. Perception of task difficulty

The first angle of looking at the data was to have people subjectively rate the children's non-verbal expressions. To investigate

whether adults can correctly interpret children's reactions to arithmetic problems and whether they do this better based on auditory and/or visual information, we carried out a perception experiment.

3.1. Method

3.1.1. Design

The experiment was a mixed design, with grade (groups 4 and 7) and level of difficulty (easy, difficult) as within-subjects variables and mode (audio, video, audio + video) as between-subjects variable. The order of the stimuli was randomly varied to prevent order-effects.

3.1.2. Participants

Fifty one adults participated in this experiment (16 male, 35 female). Their mean age was 21 years old (SD = 4). 94% of these participants were attending or had finished higher education at the time of the experiment.

3.1.3. Materials

3.1.3.1. Introduction. In the introduction participants were told they would see and/or hear (depending on condition) a number of clips from children giving answers to arithmetic problems, and that for each answer they had to indicate on a 7-point scale whether the child thought the problem was easy or difficult. To prevent them from paying attention to the outcome of the sum in audio-conditions, we told participants that the answer did not relate to the sum being easy or difficult (''every answer can be an answer to an easy (1 + 1 = 2) or difficult problem (98–96 = 2)'').

3.1.3.2. Stimuli. A data set consisting of 114 audio and/or video recordings of the children was given to all participants. Every child appears twice in the stimuli set, once when answering an easy problem (1 + 1 = 2), and once when answering a difficult problem (87–12 = 75 for group 4; 193 + 159 = 352 for group 7). We used 114 instead of all 116 stimuli, because the recordings of two children were not of good enough quality to use them for the perception experiment.

The stimuli were cut from the moment the children had seen the problem until they had given an answer. Every stimulus started with a number and a sound. We used Windows Media Player to play the films, omitting audio for the video-conditions, and video for the audio-conditions.

3.1.3.3. Answer sheet. The participants in the perception experiment rated all 114 clips on their perception of the child's experience of difficulty-level on a seven point scale from very easy to very difficult. On the last page of the answer sheet, one open question was asked in which participants could indicate on what grounds they determined their answers.

3.1.4. Procedure

First, two stimuli were presented as example, on which participants were asked to rate the child's experience of difficulty level. They could then ask for clarifications if necessary. Then the 114 experimental stimuli followed. After each stimulus, participants had 4 s to indicate the perceived level of difficulty. Then the number and sound of the next stimulus appeared automatically. After the experiment had ended participants answered the last question about the grounds on which they decided on the difficulty level and were thanked for their participation.

3.2. Results

3.2.1. Can adults interpret non-verbal cues to infer whether children are dealing with an easy or difficult arithmetic problem?

We found main effects for difficulty level, $F(1,48) = 770,17$, $p < .001$, $\eta^2 = .94$, and group, $F(1,48) = 254.04$, $p < .001$, $\eta^2 = .84$. Easy problems, $M = 2.35$, $SD = 0.07$, were indeed rated as easier than difficult problems, $M = 4.23$, $SD = 0.06$. Problems were rated differently in group 4, $M = 3.47$, $SD = 0.06$, than in group 7, $M = 3.10$, $SD = 0.06$ (see Table 1). There was also an interaction effect between difficulty level and group, $F(1,1) = 2378,68$, $p < 001$; the difference between easy and difficult problems was rated larger in group 7 than in group 4 (see Fig. 3). In other words, it appeared to be easier to contrast easy and difficult problems in group 7 than in group 4.

We found a significant effect of modality, $F(2,48) = 3548,91$, $p < .001$, $\eta^2 = .26$. Post-hoc tests revealed that when only audio was rated, children were perceived to have more difficulty with the arithmetic problems, than in the video and video + audio modalities, see Fig. 4. However, there was no interaction effect between modality and difficulty level, $F(2,48) = 1.52$, $p = .23$. It is not easier to detect easy or difficult problems when listening to speech, seeing facial expressions or both.

3.2.2. What features do adults take into account when interpreting children's affective state?

Over 94% of all participants mentioned they looked at children's expression in face and/or voice. Their explanations were mostly given in general terms, such as paying attention to 'facial expressions', or 'manner of talking'. Some participants were more specific and mentioned 'eyes', 'looking away', 'moving the lips', 'frowning', 'tone of voice'. Another feature mentioned regularly (39%) had to do with (un)certainty or hesitation. Adults either used these general terms in their answers, or they explained more specifically what they paid attention to, such as 'whether children gave their answers in a determined way, or inarticulately/ hesitantly', or 'volume of the voice'. Uncertainty or hesitation was mostly explained using voice characteristics instead of facial expressions.

Over 86% of all participants mentioned 'pause' or 'delay in answering' as a feature they based their decisions on. Only one person mentioned delay in answering as the only feature she paid attention to, all others mentioned more features.

3.3. Discussion

The results of this perception experiment show that adults can indeed interpret children's non-verbal characteristics correctly when they are answering easy or difficult arithmetic problems.

Although all participants except one reported more than just 'pause' as a feature they paid attention to, we wondered whether people could still detect task difficulty when they could not hear or see how long it took for a child to give an answer. Therefore, we carried out a second perception test, in which we cut the clips from 1 s before giving an answer until the answer was given, i.e., pause was eliminated and all clips were about the same length.

Table 1
Means and standard deviations of adults' ratings in the three modalities with visual pause.

	Video *M* (*SD*)	Audio *M* (*SD*)	Video + audio *M* (*SD*)
Group4 – easy	2.58 (0.48)	3.23 (0.38)	2.78 (0.53)
Group4 – difficult	3.96 (0.44)	4.36 (0.36)	3.93 (0.39)
Group7 – easy	1.54 (0.47)	2.25 (0.65)	1.71 (0.56)
Group7 – difficult	4.23 (0.55)	4.62 (0.48)	4.26 (0.38)

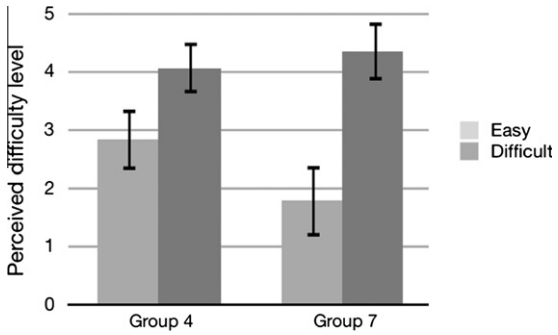

Fig. 3. Mean ratings for easy and difficult arithmetic problems in group 4 and group 7. Error bars represent standard deviation.

4. Perception of task difficulty without delay in answers

4.1. Method

4.1.1. Participants

In this experiment 52 adults participated (18 male, 34 female). Their mean age was 28 years old ($SD = 11$). 87% of these participants were attending or had finished higher education at the time of the experiment. None of them participated in the previous experiment.

4.1.2. Materials and procedure

Materials and procedure were exactly the same as in the previous perception analysis, except for the length of the clips. In the previous experiment, stimuli were cut from the moment the children had seen the problem until they had given an answer, while in this experiment, all stimuli were cut from 1 second before giving an answer, i.e., all stimuli were about the same length. Also, we used 112 clips, because 2 clips were too short to cut.

4.2. Results

When pause is excluded from the material, adults can still interpret whether children are solving easy or difficult arithmetic problems. We found main effects again for difficulty level, $F(1,49) = 5.04$, $p < .05$, $\eta^2 = .09$, and group, $F(1,49) = 144.92$, $p < .001$, $\eta^2 = .74$, albeit the effect of difficulty level is much weaker. We also, again, found an interaction effect between difficulty level and group $F(1,49) = 57.45$, $p < .001$, $\eta^2 = .54$.

There was no significant effect of modality alone anymore, $F < 1$. Surprisingly however, we found an interaction effect between difficulty level and modality, $F(2,49) = 10.18$, $p < .001$, $\eta^2 = .29$. Fig. 5 shows that in all modalities, it becomes harder to distinguish between the easy and difficult problems. However, in the video and video + audio modality, the easy problems are still rated easier than the difficult problems, while in the audio condition, the easy problems in group 4 are rated as more difficult than the difficult problems.

When asked what cues adults used to decide on difficulty, more detail was given in their answers, probably because they could not rely on delay in answering anymore. The same categories can be applied as before: participants talking about expression in face and voice, mentioning 'uncertainty' in particular. 'Showing pride' and 'showing boredom' were also mentioned as general features. More specifically, they mentioned 'pitch', 'stuttering', 'loudness', 'intonation', and 'thinking aloud' for voice, and 'smiling', 'looking away', 'movement of the eyes', 'staring', and 'position of the mouth' for facial expressions.

4.3. Discussion

When pause is excluded from the material, adults can still interpret whether children are solving easy or difficult arithmetic problems. However, it is harder to distinguish the easy and difficult problems, especially in the audio-condition, suggesting that visual features are the most important ones. Taken together, the results of the two perception experiments suggest that non-verbal expressions vary with age. To further investigate which non-verbal cues children show and whether these are related to the self-reports in the perception experiments, we analyzed the facial expressions of the children in the 114 fragments by coding them. Auditory cues were not included in the analyses, since the results of the two perception tests suggest that visual cues are more important than auditory ones for problem detection.

5. Analysis of facial expressions

The second angle of looking at the data was to objectively rate the children's facial expressions. Five expressions were chosen based on what our participants reported, and on previous research (Swerts & Krahmer, 2005).

5.1. Method

The 114 fragments of children's reactions to arithmetic problems were coded on five facial expressions: (1) smiling, (2) gaze, (3) frowning, (4) 'funny face' (a marked facial expression, which diverts from a neutral expression), and (5) visual pause (Swerts &

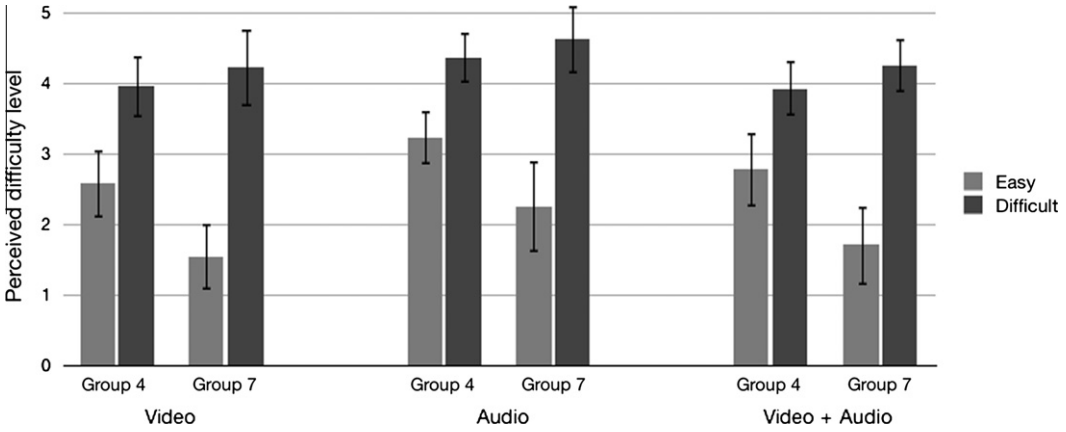

Fig. 4. Mean score of easy and difficult arithmetic problems in three conditions when pause is included. Error bars represent standard deviation.

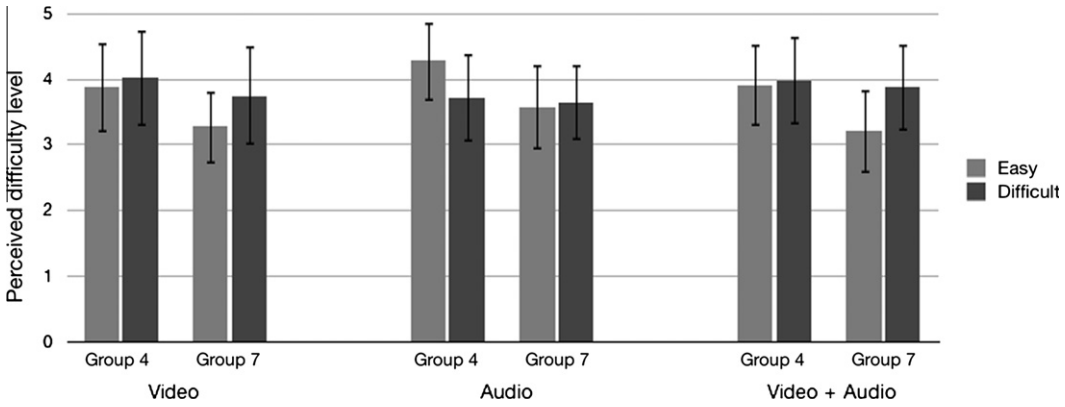

Fig. 5. Mean score of easy and difficult arithmetic problems in three conditions when pause is excluded. Error bars represent standard deviation.

Table 2
Frequencies of shown facial expressions and pause in easy and difficult arithmetic problems in two groups.

	Easy		Difficult	
	Group 4	Group 7	Group 4	Group 7
Smiling				
Yes	5	7	2	5
No	23	22	26	24
Gaze				
Yes	0	0	11	9
No	28	29	17	20
Frowning				
Yes	0	0	3	10
No	28	29	25	19
Funny face				
Yes	2	0	8	11
No	26	29	20	18
Visual pause				
Yes	4	0	22	28
No	24	29	6	1

Krahmer, 2005). Verbal cues were not included in the analysis. The features are loosely based on some of the Action Units (AUs) described by Ekman and Friesen (1978) to distinguish facial expressions and the facial muscles involved. Of the visual features under consideration here, smiling is related to AUs 12 and 13 and gaze to AUs 61–64. Frowning is related to AUs 1 and 2. Funny faces typically consist of a combination of AUs such as lip corner depression (AU 15), lip stretching (AU 20) or lip pressing (AU 24), combined with eye widening (AU 5) and possibly brow movement as well. Representative examples of these facial expressions are displayed in Fig. 2.

For every fragment, the five expressions are scored as present (=1) or not present (=0). Each reaction to an arithmetic problem could thus score a minimum of 0 facial expressions and a maximum of 5 facial expressions.

5.2. Results

Table 2 shows the frequencies (114 fragments) of shown facial expressions, split by difficulty level of the arithmetic problem and group level.

On average, children show significantly fewer expressions when they are faced with an easy problem ($M = 0.32$; $SD = 0.51$) than when faced with a difficult problem ($M = 1.91$, $SD = 1.01$), $t(114) = -12.41$, $p < .001$. Smiling occurs both when facing easy and difficult problems, $t(114) = 1.25$, $p = .21$, while all other facial expressions occur significantly more often when a problem is difficult than when it is easy ($p < .001$).

There is no significant difference in amount of expressions shown between group 4 and group 7, $t(114) = -.51$, $p = .61$.

5.3. Discussion

The results of the analysis of facial expressions confirm that the facial expressions as mentioned in the self-ratings by participants indeed occur, and that they occur more in difficult problems than in easy problems. We can conclude that it is possible to derive difficulty level from facial expressions, and that adults are able to correctly 'read' these expressions. To investigate whether a computer can also detect the difficulty level from non-verbal cues, we trained a computer on data previously gathered.

6. Automatic detection of difficulty level

The third angle of looking at the data was to try to automatically rate the children's facial expressions. State-of-the-art facial expression recognition methods may be able to capture nonverbal information, thereby confirming or complementing the human assessments. In order to address whether state-of-the-art facial expression recognition methods are able to capture the nonverbal expressions of the children, we performed a computational analysis of our fragments using the Active Appearance Models (AAMs) method (Cootes, Edwards, & Taylor, 2001). This method has been successfully applied to facial expression recognition (Cohn, 2010) and requires the manual specification of a grid of landmarks for a few representative frames of the video fragments. For facial expression analysis, the landmarks are placed at facial locations whose positions and appearances are of relevance for expressions. Fig. 6 is an illustration of such a grid superimposed on a single frame of a fragment. The dots represent the landmarks and the lines connecting the dots form the grid. Once representative grids are created, the AAM method uses these to derive a model that captures the facial movement. With this model a grid can be automatically created for the rest of the frames.

6.1. Thin-slice analysis of facial expressions

Our perception experiment with adults showed that pause information is a strong cue for problems, so including these in the automatic analysis would be easy but arguably somewhat uninformative. Here we want to see whether automatic detection is also possible on very brief fragments. Inspired by the notion of *thin slicing* (Ambady & Rosenthal, 1992; Gladwell, 2005), we decided to restrict our computational analysis to the first second (25 frames) of each fragment. The goal of our analysis was to determine to what extent the head movements that immediately follow

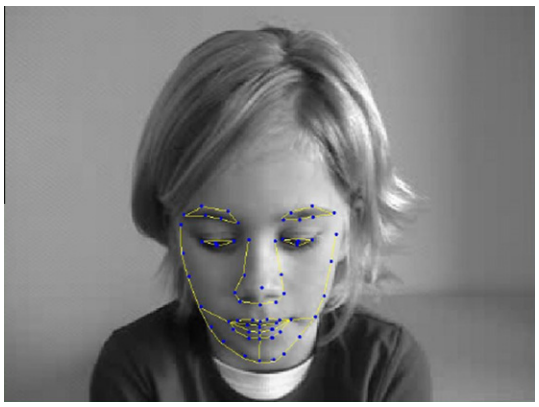

Fig. 6. Illustration of a manually applied grid (lines) of landmarks (dots) superimposed on a single frame in one of our fragments.

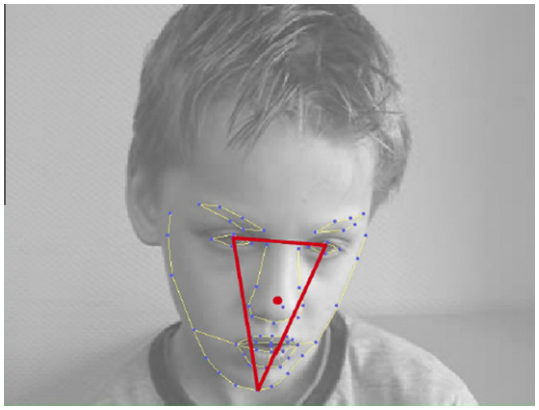

Fig. 7. Illustration of the subset of three landmarks (the corner points of the triangle) used for computing facial movements. The large dot indicates the center-of-mass of the three landmarks.

the presentation of the arithmetic problem, reveal signs of the level of difficulty of the problem. In order to perform the analysis, we used the AAM method to fit a grid of 66 landmarks (see Fig. 6) to the faces of all participants. To this end we used a Matlab implementation of the AAM method (Van der Maaten & Hendriks, 2010). At the expense of the time-consuming careful manual positioning of landmarks for a few frames, the resulting fit was quite good; it compares favorably to errors reported in the literature (Joosten, 2011).

For the measurement of head movements, we used a very small subset of fitted landmarks, i.e., those located at the two eyes, the tip of the nose, and at the chin. From this subset of landmarks, we computed the coordinates of the center-of-mass as an estimate of the three-dimensional pose of the head. Fig. 7 displays a superimposed grid and the associated center-of-mass (large dot) of three landmarks (the corner points of the triangle). The dynamics of the head movements were captured by analyzing the movements of the center-of-mass in all pairs of subsequent frames. For each 25-frame fragment, we obtained 24 pairs of coordinates constituting *head-motion vectors*. The polar coordinates of the head-motion vectors were used to represent the angle and magnitude of the head movements.

We employed a non-parametric histogram classifier to predict the type of arithmetic questions (easy versus difficult) from a single 25-frame fragment. The non-parametric histogram classifier relies on a two-dimensional histogram as a density estimate of the distribution of the polar coordinates of the head-motion vectors. The classifier is evaluated in a leaving-one-out cross-validation procedure as follows. Given the data set of 110 video fragments (two were taken out because the faces could not be detected well), one fragment is defined as the test fragment and the remaining fragments as training fragments, about half of which belongs to the *easy* class and the other half to the *difficult* class. Separate two-dimensional (angle and magnitude) histograms are created for each class using the training fragments, and both histograms are normalized. The two-dimensional histogram of the test fragment is multiplied with both training histograms and the results are summed. If the sum of the multiplication with the *easy* histogram is larger than that of the *difficult* histogram, the test fragment is assigned to the *easy* class, otherwise it is assigned to the *difficult* class. This procedure is repeated 110 times in such a way that each fragment becomes a test fragment once. For each repetition, the classification of the test fragment can be wrong or right. The percentage of correctly classified test fragments is the prediction performance.

6.2. Results of the thin-slice analysis

Differences in head movements for the easy and the hard questions were already visible in the raw coordinates of the center-of-masses. Fig. 8 shows the center-of-mass coordinates during the first 25 frames for the easy arithmetic problems (top row) and the difficult ones (bottom row) as sequences of points. The left column displays the sequences of points for all participants, the middle and right columns show the results for the fourth and seventh group participants, respectively. From the overall pattern of these plots a clear difference in prevailing orientation of the points is visible. Whereas in the top row (easy problems) the sequences of points are oriented mainly vertically, in the bottom row (difficult problems) the sequences of points are oriented more in the diagonal direction.

The classifier achieved a prediction performance of 71% as determined using the leaving-one-out cross-validation procedure. This implies that the perceived difficulty level of the arithmetic problem can already be predicted from the first second of head movement with an accuracy of 71%. The result obtained differs significantly from chance level (50%) according to a binomial test, yielding a p-value of $p = 4.1062 \times 10^{-6}$. When training our classifier on fourth or seventh group fragments only, we obtained performances of 67% and 63%, respectively; these scores are presumable somewhat lower (though still significant) due the availability of fewer data.

As is evident from the plots in Fig. 8, our prediction performance is due to a difference in orientation of head movements during the first second after being confronted with the arithmetic problem. We carefully examined the fragments to detect these head movements. This turned out to be quite hard. Nevertheless, we were able to identify some illustrative examples, one of which is shown in Fig. 9. The panels in this figure show the begin frames ((a) and (c)) and end frames ((b) and (d)) of a head-movement sequence when the participant is confronted with an easy ((a) and (b)) or hard problem ((c) and (d)). In each panel, the left part is the original frame and the right part shows the AAM-fitted grid of landmarks. In (b) and (d), the right parts show the head-movement traces (red lines). The vertical head movement associated with an easy problem is clearly visible in (a) and (b), whereas the movement in the diagonal orientation, characteristic for hard problems, is present in (c) and (d).

6.3. Discussion

Our computational analysis revealed that it is possible to automatically identify facial expressions related to the perceived level of difficulty of arithmetic problems. Our method correctly

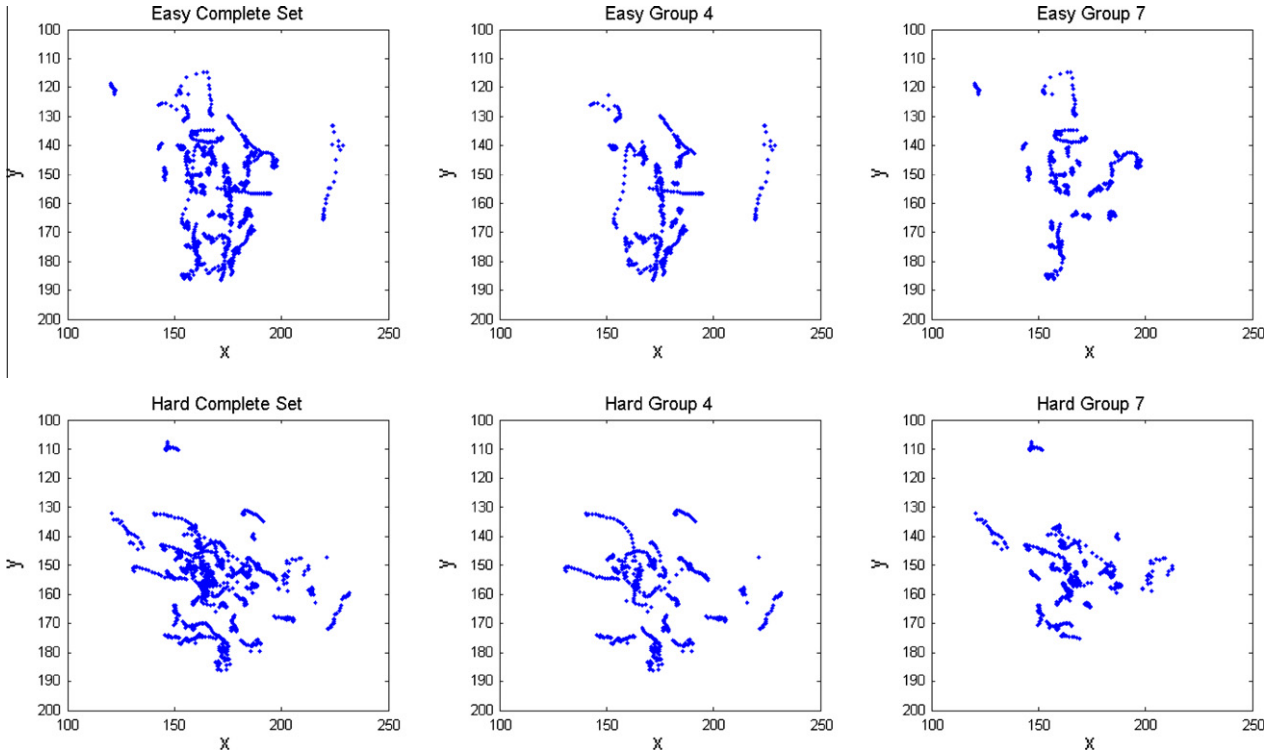

Fig. 8. Plots of the center-of-mass coordinates during the first 25 frames (x and y are screen coordinates in pixels) for the easy problems (top row) and difficult problems (bottom row). The left column shows the coordinates for all participants, the middle and right columns those for the fourth and seventh group children.

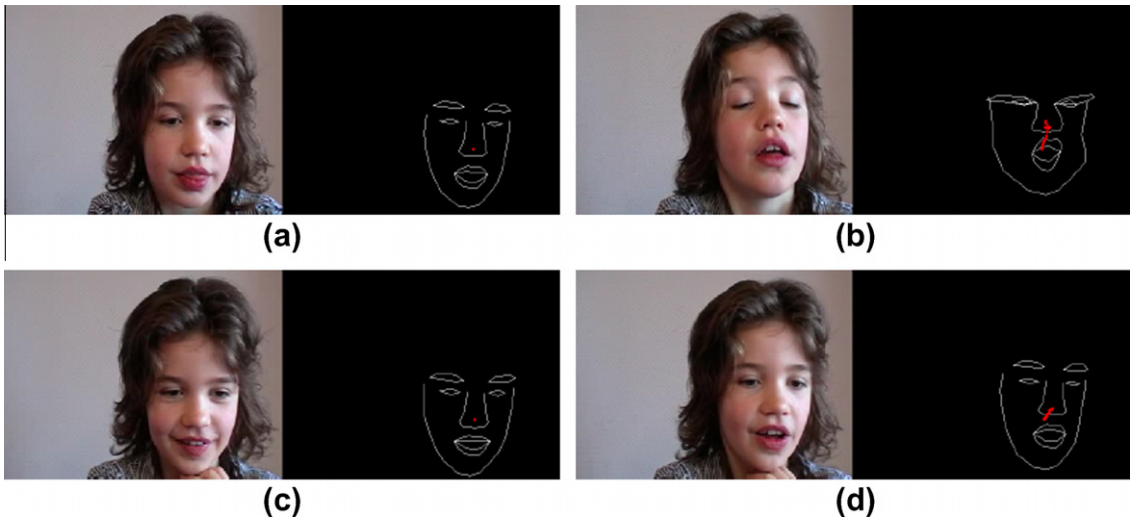

Fig. 9. Illustration of the behavioral pattern discovered through our computational analysis of the 1-s thin slice. (a) and (b) show the start and end frames of a vertical head movement displayed by the participant after being confronted with an easy problem (c) and (d) are the start and end frames for a difficult problem. Each panel shows the original frame on the left and the fitted model (grid of landmarks) on the right. In (b) and (c) the lines represent the trace of head movement.

identified whether children thought a problem was easy or difficult in 71% of all cases in the first second after viewing the problem. The analyses are based on general head movements, and not on the specific facial features mentioned by human raters.

7. General discussion

When engaged in school tasks, children may be in an affective or cognitive state that helps or hinders their task execution. They might think a problem is easy or difficult and behave accordingly. Our research has shown, first, that such states (more specifically experiencing an arithmetic problem to be easy or difficult) are shown in children's non-verbal behavior, and second, that adults are able to correctly interpret this non-verbal behavior. It is conceivable that speakers unconsciously use such cues to signal frustration or boredom, so that addressees (such as teachers) can adapt their behavior accordingly.

We found that adults can correctly interpret non-verbal cues both when looking at children's faces, and when listening to their voice or a combination of these two modalities. Pause was one of the strongest cues, but even when pause was omitted in the stimuli, adults were still able to determine whether a child found the problem easy or difficult based on just the answer. We found that

visual cues were more useful for participants, and auditory cues appeared to be of little or no added value in our findings. The combination of both modalities seems to follow the line of the video-modality only. This is consistent with the finding that perception of emotions is stronger in visual than in auditory stimuli. For example, when people have to judge the emotion of incongruent stimuli, such as a happy face with a sad voice, the visual cues have a stronger influence (e.g., Hess, Kappas, & Scherer, 1988; Massaro & Egan, 1996). Dijkstra et al. (2006) also found that in detecting uncertainty when answering questions, facial expressions had a much stronger effect than auditory fillers such as 'uhm' and intonation. It should be noted that the auditory cues in our experiments were not very informative; children give an answer that consists only of a number, not a complete sentence, so there is relatively little room for prosodic cues. To infer affective state from such a short cue must be very hard. Although Dijkstra et al. found that auditory fillers did not have a strong effect compared to facial cues, the difference between the audio-conditions with and without pause may be explained by hearing fillers such as 'uhm' or sighs in the 'with pause' condition on top of just hearing the answer without pause.

In general, participants had more trouble distinguishing easy from difficult arithmetic problems in group 4 than in group 7. This is somewhat surprising, because based on for instance Thompson (1994) and Doherty-Sneddon and Kent (1996), we could have expected younger children to be more expressive than older children. However, our findings are consistent with findings by e.g., Shadid, Krahmer, and Swerts (2008) that 12 year old (Pakistani and Dutch) children were more expressive playing a game than 8 year olds. Distinguishing different types of emotions as in Adolphs (2002) article, such as social emotions, motivational stated and moods, may be appropriate here.

The non-verbal cues participants mainly focused on were smile, gaze, tone of voice and pause when applicable. Frowning was less often mentioned, and funny face was not mentioned as such, although it is not clear whether participants implicitly used this feature. The results of the analysis of facial expressions confirms that these expressions indeed occur, and that they occur more in difficult problems than in easy problems.

The analysis of facial expressions does not show a difference between group 4 and group 7, while our perception experiments showed that it was easier for the participants to distinguish easy from difficult problems in the older age group. The age difference may be a gradual one, making it harder to demonstrate by counting separate features. Moreover, it may well be possible that the features we analyzed were not the only ones adults paid attention to, or rather, that it is a combination of features that is important. In fact, people and computers used different cues to interpret children's facial expressions when answering an arithmetic problem. People mainly referred to specific facial cues. The computer used head movements to infer children's state. The vertical and diagonal head movements seem to be related to the approach-avoidance dimension used in emotion literature. Frijda and Tcherkassof (1997) argued that facial expressions appear in a context of head and body orientation. For example, a frightened face is often accompanied by withdrawal movements of head and shoulders. Similarly, the vertical head movements in the easy problems are related to approach, and the diagonal movements in the difficult problems to avoidance.

The results of our computational analysis are promising for future Adaptive Tutoring Systems. They revealed a behavioral pattern (i.e., vertically versus diagonally oriented head movements) in a thin slice of 1 s after presentation of the arithmetic problem that was not noted by any of our human raters of the fragments. Even when knowing where to look for it is still hard to detect these behavioral patterns. Apparently, the behavioral pattern is quite subtle and can only be revealed through computational analysis based on a facial expression model and a classifier.

Although our computational analysis may readily be extended by incorporating more landmarks to analyze more subtle expressions and by using more sophisticated classifiers (or by including a 'cheap' future, such as pause), we are encouraged by our result because, theoretically, it suggests (1) that computational analysis may lead to the discovery of hitherto unnoticed non-verbal behavior patterns, and practically (2) that it is possible to automatically estimate the experienced difficulty from facial expressions, which is of relevance for automatic tutoring systems.

Although technically possible, more research is needed before ATSs can be implemented in schools. We first need to find out how a system should *react* to a child's state. For example, when a child perceives an arithmetic problem as mildly difficult, it might be good to continue showing problems of the same level, while a frustrated child that believes he can never solve these problems would be motivated by receiving arithmetic problems of lower difficulty level (see Kapoor et al., 2001). Also, further research could address whether our results are generalizable to domains other than arithmetic problems, so students can learn with Adaptive Tutoring Systems on more fuzzy tasks for example.

8. Conclusion

In this paper we asked the question whether children's non-verbal cues can be used to determine whether children have performance difficulties with arithmetic problems. Using a new experimental paradigm (based on "brain training" software) we collected the reactions from primary school children to easy and difficult arithmetic problems. We found that adults can successfully interpret the difficulty level of these problems based on non-verbal cues of the children, especially when they could see their faces. An automatic, "thin slice" analysis revealed that it is possible to automatically predict, above chance, whether a problem was easy or difficult. This finding has potentially important consequences for Adaptive Tutoring Systems that can detect a child's state and adapt the difficult of the learning material accordingly, which would make it possible to pertain to the needs of every individual child even when learning together in classrooms.

Acknowledgements

We would like to thank Joost Driessen, Lucinda Martens, and Hans Westerbeek for their help in preparation of the stimuli and collection of the data.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*, 21–62.

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*, 127–145.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.

Back, E., Jordan, T. R., & Thomas, S. M. (2009). The recognition of mental states from dynamic and static facial expressions. *Visual Cognition, 17*, 1271–1286.

Barkhuysen, P., Krahmer, E., & Swerts, M. (2005). Problem detection in human–machine interactions based on facial expressions of users. *Speech Communication, 45*, 343–359.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to observers about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User modeling and user adapted interaction, 6*, 87–129.

Cohn, J. F. (2010). Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Processing Magazine, 27*, 128–133.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*, 681–685.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2007). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition & Emotion, 22*, 777–788.

Craig, S. D., Graesser, A. C., Sullens, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*, 241–250.

De Silva, L. C., Miyasato, T., Nakatsu, R. (1997). Facial emotions recognition using multi-modal information. In *Proceedings of IEEE International Conference on information, communication and signal processing*, Singapore, pp. 397–401.

De Vos, T. (1994). *Tempo Test Rekenen [Arithmetic Speed Test]. Handleiding Tempo-Test-Rekenen* (2nd ed.). Lisse, The Netherlands: Swets Test Publishers.

Dijkstra, C., Krahmer. E., Swerts, M. (2006). Manipulating Uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. Paper presented at the Speech Prosody conference, Dresden, Germany.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R., Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. In *Proceedings of the international conference on information, communications and signal processing* (pp. 397–401). USA.

Doherty-Sneddon, G., & Kent, G. (1996). Visual signals and the communication abilities of children. *Journal of Child Psychology & Psychiatry, 37*, 949–959.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding scheme*. Palo Alto, CA: Consulting Psychologists' Press.

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranach et al. (Eds.), *Human Ethology*. Cambridge: Cambridge University Press.

Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). Cambridge: Cambridge University Press.

Gladwell, M. (2005). *Blink*. New York: Little, Brown and Company.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Graf, S., Liu, T-C., Kinshuk Chen, N-S., & Yang, S. J. H. (2009). Learning styles and cognitive traits – Their relationship and its benefits in web-based educational systems. *Computers in Human Behavior, 25*, 1280–1289.

Hess, U., Kappas, A., & Scherer, K. (1988). Multichannel communication of emotion: Synthetic signal production. In K. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 161–182). Hillsdale, NJ: Erlbaum.

Huang, T. S., Chen, L. S., Tao, H., Miyasato, T., Nakatsu, R. (1998). Bimodal emotion recognition by man and machine, *ATR workshop on virtual communication environment*. Japan.

Joosten, B. (2011). Facial expression recognition. *Towards digital support for behavioral scientists*. Unpublished master's thesis, Tilburg University, Tilburg, The Netherlands.

Kapoor, A., Mota, S., Picard, R. W. (2001). Towards a learning companion that recognizes affect. In *AAAI fall symposium*, November, 2001.

Knapp, M. L., & Hall, J. A. (2006). *Nonverbal communication in human interaction* (6th ed.). Wadworth: Wadworth publishers.

Krahmer, E. J., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech, 48*, 29–54.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., et al. (2011). The Computer Expression Recognition Toolbox (CERT). In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 298–305), Santa Barbara, USA.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgments. *International Journal of Behavioral Development, 26*, 327–333.

Massaro, D., & Egan, P. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review, 3*, 215–221.

Read, J., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations. In M. Bekker et al. (Eds.), *Interaction design and children*. Eindhoven, The Netherlands: Shaker Publishing.

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). How do you know that I don't understand? A look at the future of intelligent tutoring systems. *Computers in Human Behavior, 24*, 1342–1363.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227–256.

Shadid, S., Krahmer, E., Swerts, M. (2008). Alone or together: Exploring the effect of physical co-presence on the emotional expressions of game playing children across cultures. In: P. Markopoulos et al. (Eds.), Fun and Games. *Lecture notes in computer science* (pp. 94–105), 5294. Berlin: Springer-Verlag.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25–38.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*, 81–94.

't Hart, J., Collier, R., Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Thompson, R. (1994). Emotion regulation: A theme in search of definition. In: N. Fox (Ed.), The development of emotion regulation: biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*. (pp. 25–52).

Van der Maaten, L., Hendriks, E. (2010). Capturing appearance variation in active appearance models. In *Proceedings of the 23rd IEEE computer vision and pattern recognition workshops* (pp. 34–41).

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Yoshitomi, Y., Kim, S.-I., Kawano, T., Kitazoe, T. (2000), Effect of sensor fusion for recognition of emotional states, using voice, face image, and thermal image of face. In *Proceedings of 9th IEEE International Workshop on Robot and Human Interactive, Communication* (pp. 178–183).