

On the assessment of audiovisual cues to speaker confidence by preteens with typical development (TD) and atypical development (AD)

Swerts, M.G.J.; de Bie, C.

Published in:

Proceedings of the 13th International conference of the international Speech Communication Association (Interspeech)

Document version:

Publisher's PDF, also known as Version of record

Publication date:

2012

[Link to publication](#)

Citation for published version (APA):

Swerts, M. G. J., & de Bie, C. (2012). On the assessment of audiovisual cues to speaker confidence by preteens with typical development (TD) and atypical development (AD). In *Proceedings of the 13th International conference of the international Speech Communication Association (Interspeech)* (pp. 1314-1317). Red Hook, U.S.A.: Curran Associates, Inc..

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

On the assessment of audiovisual cues to speaker confidence by preteens with typical development (TD) and a-typical development (AD)

Marc Swerts and Cees de Bie

Tilburg University, School of Humanities, TiCC research center, Tilburg (The Netherlands)

m.g.j.swerts@uvt.nl

Abstract

This paper looks into how preteens with autism (Asperger, PDD-NOS) compare to healthy controls (matched in terms of age, IQ and educational level) in the way they interpret audiovisual expressions produced by adult or child speakers. In previous research, we had recorded utterances from those groups of speakers as they were responding to easy and difficult questions in a quiz-like experiment, so that they were not always equally confident about the correctness of a given answer. The task given to the preteens in the current study was to judge how certain a speaker appeared in his/her response to a question, where they could base such judgments both on auditory and visual properties of the speakers. Results reveal that all groups of preteens are able to estimate a speaker's confidence level on the basis of audiovisual properties, albeit that the preteens diagnosed with PDD-NOS performed significantly worse at this than the other two groups. Moreover, in line with previous results, participants found the data coming from child speakers harder to judge than those produced by adult speakers.

Index Terms: nonverbal communication, prosody, cues to speaker confidence, autism

1. Introduction

It has repeatedly been argued that one of the defining characteristics of children with autistic development (AD) is that they tend to be more deficient in their social functioning than typically developing (TD) children (Kanner 1943; Baron-Cohen 1995). This would appear from the fact that they find it comparatively more difficult than healthy controls to interpret facial expressions, and to integrate such nonverbal cues with the spoken content of incoming messages. In line with this observation, people with autism have often been claimed to be particularly impaired in how they recognize emotions in others. However, others have questioned this claim and have argued that the problem is more nuanced, in particular with studies showing that high-functioning people with autism “do not have a specific deficit in affect recognition that differentiates them from nonautistic people of similar developmental level” (Loveland et al. 1997). Instead, Baker et

al. (2009) suggest that the problems reported before are less due to autism, but more to more general cognitive impairments.

What complicates the discussion regarding the extent to which people with autism do or do not differ from healthy controls, is the observation that many previous studies in this area are based on data that are arguably not ecologically valid. Notwithstanding a few exceptions, much research in the past has been based on stimuli consisting of still images, like pictures or drawings of faces. In most of their interactions with others, however, people have to process “fleeting changes” in facial expressions (Adolphs 2002). Moreover, the still images in prior research were mostly presented to participants without any accompanying sound, which, again, is different from a whole range of contexts where people need to process face and voice at the same time. This is a potentially crucial factor when dealing with AD participants because of reported findings that people with AD tend to find it harder than healthy controls to integrate multimodal input, such as voice and face (Davis et al. 2006; Dunn et al. 2002). An additional peculiar aspect of previous data is that they often tend to be acted versions of basic emotions (like anger or fear), which again may yield stimuli that are not representative of expressions under more naturalistic conditions. In fact, from previous work (Wiltling, Krahmer & Swerts 2006), we know that acted emotions, even though they may not be “felt” as such, tend to be perceived as more exaggerated and more stereotypical than emotions that are spontaneously elicited.

2. Cues to a speaker's confidence level

Given that most of the research into the nonverbal skills of people with AD have focused on how they deal with nonverbal correlates of basic emotions, the current study wants to explore how they handle what has been termed social emotions, using ecologically valid stimuli. In particular, we will focus on cues to uncertainty, which has been argued to be a typical social emotion (Swerts & Krahmer 2005). It has been shown that such kinds of social emotions can be culturally quite different, and need to be learned naturally through interactions with others. In our study, we will specifically focus on high-



Figure 1: *Stills of facial expressions low FOK expressions by an adult and a child speaker*

functioning people with AD, as there has been a discussion as to whether they do or do not differ in their non-verbal skills from healthy controls.

In previous research, we found that speakers indeed tend to give cues to uncertainty. Following the Feeling-of-Knowing (FOK) paradigm, originally introduced by Hart (see e.g. Brennan & Williams 1995), we elicited certain and uncertain responses from speakers through a quiz-like experiment (see Swerts & Krahmer 2005; Krahmer & Swerts 2006). In the experiment, participants, while being video-taped, are instructed to give answers to a series of different questions (e.g. What is the capital of Switzerland? How many degrees in a circle?, ...) that vary in degree of difficulty, depending on a speaker's interest or prior knowledge. Typically, in the experiment, speakers will not always be equally confident about the correctness of their answer to a specific question. In the experiment, we also ask speakers after the quiz to indicate on a scale how confident they are that they would recognize the correct answer to a question in a multiple-choice test. This score is known as the Feeling-of-Knowing (FOK) score. We conducted this experiment both with adult speakers and with children in the age of 7-8. It turned out that they signal their confidence level nonverbally, both with visual and auditory features. When being uncertain, speakers have a tendency to produce a puzzled look, frown their eyebrows or turn away their gaze, and also are more likely to produce an answer after some delay, with filled pauses and a question intonation. Some representative stills of low FOK expressions produced by an adult and child speaker are shown in Figure 1.

Interestingly, it turned out that adults are more expressive than children in how they use such features to mark their level of uncertainty. In a perception experiment in which participants were asked to rate the confidence level in audiovisual recordings of speakers, the so-called Feeling-of-Another's-Knowing (FOAK), we found that adult and child judges can more easily distinguish uncertain from certain responses in data from adults than in those from children. In addition, children perform more poorly as judges than adults when having to assess

a speaker's confidence level (Krahmer & Swerts 2006). Both these production and perception results are in line with the assumption that the display and interpretation of (un)certainty is a social skill that children need to acquire as part of their development. As a matter of fact, older children in the age of 12 tend to show their confidence level clearer than 8-year old children (Visser et al. 2010).

So given these results, our current study investigates how preteens with typical development or with autism would assess speakers' cues to confidence level. Since we know that autism represents a whole range of cognitive and communicative impairments, we decided to look at the performance of people with two kinds of autism-related characteristics, i.e., PDD-NOS and Asperger, who are matched as much as possible to a group of healthy controls.

3. Perception study

3.1. Method

3.1.1. Stimuli

The stimuli used for the current perception experiment are the same ones as those used in the perception experiment discussed in Swerts and Krahmer (2005) and in Krahmer and Swerts (2006). In particular, they consist of 60 responses of various speakers who had previously participated in a feeling-of-knowing study, where they had to answer questions that differed in level of difficulty. Half of those stimuli came from answers produced by adults, and half by 8-year old children. Half of the stimuli were certain responses (as indicated by feeling-of-knowing scores), and half were uncertain. More details can be found in Krahmer and Swerts (2006).

3.1.2. Procedure

The testing procedure was organised as a group experiment in which participants had to rate a sequence of videoclips, where they had access to both video and sound, though each participant had to make his/her judgment on an individual answer sheet. The actual experiment was preceded by a short session with 4 clips (different from the ones shown in the actual experiment) to enable participants to familiarize themselves with the kinds of video clips. As in previous studies, the participants were asked to give FOAK-scores.

3.1.3. Participants

Our experimental group of participants consisted of 56 preteens (48 male; 8 female) who had a CITO score¹ of at least 540, a total IQ score of 110 on WISC-III, and a recommendation of their primary school to attend

¹A CITO score is used in the Dutch school system to advice children in their final year of primary school on what school to attend later on.

Table 1: Mean FOAK and difference scores for high-FOK and low-FOK answers by non-autistic preteens, and preteens with Asperger or PDD-NOS.

Preteens	High FOK	Low FOK	Δ -score
Control	5.033	2.599	2.434
Asperger	5.109	2.823	2.286
PDD-NOS	4.880	2.872	2.008

HAVO/VWO (a specific educational level in the Dutch school system). They were all officially diagnosed with autism (high-functioning), using criteria of DSM-IV-TR. Of this group, 25 children had been diagnosed with Asperger, and 31 children with PDD-NOS. Their average age was 12.4 (min: 11 years old; max: 17 years old). In addition, we collected data from a control group of 55 healthy children (28 male; 27 female) who had received a CITO score of 538 as a minimal requirement to register for the school, and a recommendation of their primary school to attend HAVO/VWO. Their average age was 13.5 (min: 12 years old; max: 14 years old). So while the experimental and control groups were balanced in terms of IQ, educational level and age, there was a large difference in the relative representation of both genders, which is statistically significant between both populations ($\chi^2 = 15.57, p < .001$).

3.2. Results

The data were analysed with a repeated measures anova with answer (2 levels: high-FOK and low-FOK) and speaker (2 levels: adult vs child) as within-subject factors, type of judge as between-subject factor (3 levels: non-autistic, Asperger and PDD-NOS) and the average FOAK score per judge as dependent variable. The analysis revealed a main effect of answer ($F_{(1,108)} = 2078.746, p < .001, \eta_p^2 = .951$), with high-FOK answers receiving significantly higher FOAK scores than the low-FOK answers (high-FOK = 5.006; low-FOK = 2.746). There was no main effect of judge or speaker.

In addition, the two-way interaction between answer and judge also turned out to be significant ($F_{(2,108)} = 7.389, p < .001, \eta_p^2 = .120$). The corresponding average values are given in Table 1, which shows that the difference between low-FOK and high-FOK answers is somewhat bigger for the scores of the control group, and smaller for the judges diagnosed with PDD-NOS. A one-way anova with type of judge as between-subject factor and the difference scores between high and low FOK answers as dependent variable revealed a significant main effect of judge ($F_{(2,108)} = 5.881, p < .01, \eta_p^2 = .098$), where posthoc pairwise comparisons using the Bonferroni method showed that the scores produced by the healthy controls were significantly different from those of the PDD-NOS participants, with the remaining compar-

Table 2: Mean FOAK and difference scores for high-FOK and low-FOK answers produced by adult and child speakers.

Speakers	High FOK	Low FOK	Δ -score
Adults	5.334	2.504	2.830
Children	4.678	3.025	1.653

isons not being significantly different from each other. Inspection of the values in Table 1 also reveals that the differences between the scores are mainly due to the fact that the low-FOK scores are somewhat higher for the PDD-NOS and Asperger participants. In addition, there was a significant 2-way interaction between speaker and answer ($F_{(1,108)} = 284.917, p < .001, \eta_p^2 = .725$). As shown in Table 2, the difference between low-FOK and high-FOK answers is not perceived as large as similar answers produced by adult speakers. All other interactions were not significant.

3.3. Discussion

The experiment described above has shown some differences between the different groups of preteens that we analysed in terms of how they judge the level of uncertainty in others. Interestingly, the main difference appears to be one between preteens diagnosed with PDD-NOS and preteens with typical development, in that the latter are significantly better in distinguishing highly confident from less confident speaker responses. The children diagnosed with Asperger did not perform more poorly than the healthy controls. This suggests that it is important to take the specific forms of autism into account when making claims about deficits in nonverbal behaviour. The fact that children with PDD-NOS were less accurate in their FOAK scores could be related to the fact that these children reportedly have more difficulties to concentrate and keep their focus, so that subtle cues in the auditory or visual characteristics of the recorded speakers may more easily escape their attention. More research is needed here. In any case, both the healthy controls and the children with Asperger or PDD-NOS found it harder to assess the cues in data from children than in those from adults, which is in line with previous findings that children are less explicit about their confidence level than adults. Finally, it is interesting to see that there is a tendency both for preteens with Asperger and PDD-NOS to give comparatively higher FOAK scores for the low confidence answers than the healthy controls. This does suggest that the autistic preteens have a harder time to interpret audiovisual expressions that signal uncertainty, in that sense ignoring such cues as qualifiers of the information a speaker provides. This would be in line with the fact that people with AD find it more difficult to understand irony or figure of speech, and interpret messages

more literally (Mackay & Shaw 2004).

4. General discussion

In sum, this study has looked at how groups of preteens with typical or atypical development estimate a speaker's confidence level on the basis of audiovisual properties, where we observed that healthy controls found it easier to separate certain from uncertain responses than high-functioning preteens diagnosed with PDD-NOS. In the future, we plan to extend this kind of research to the analysis of other forms of pragmatically relevant usages of nonverbal communication, and also focus on a wider range of people with autism.

Participants in our study were presented with clips in which they had access both to visual and auditory features of the speakers whose responses had to be judged in terms of level of uncertainty. It may also be interesting to run the same experiment in audio-only and video-only format. It could be that preteens with autism may have comparatively more problems when they are exposed to input coming from two modalities. Indeed, previous work has brought to light that such simultaneous presentation can be confusing for some participants with autism, so that it would be interesting to see whether their performance increases if they can concentrate on either the visual or auditory information. And it would be interesting to compare such results with those of healthy controls who have previously been shown to get better performances if they have access to multiple modalities, rather than only one (Swerts and Kraemer 2005). Moreover, it has been argued that some types of people diagnosed with autism are especially sensitive to visual cues (compared to auditory ones). Given such previous findings, it would be worthwhile to explore whether their judgments of video-only materials differ significantly from those in which only auditory cues are available.

We have so far looked at high-functioning people with autism, who attend a specialised high-schools. Obviously, it would be worthwhile to explore how their results compare to children or preteens with more severe forms of autism. If it turns out that there are (subtle) differences between these various populations, one could consider including such datasets in diagnostic procedures that try to establish the degree of autism, or in procedures to improve people's communicative skills in daily interactions. Obviously, while the judgments tasks discussed in this paper heavily rely on the metacognitive skills of participants, in treatment and diagnosis such tasks would have to be supplemented with more functional tasks in which people are trained to act upon cues regarding the mental states of other people, including their confidence level.

And finally, we have only looked at the perceptive skills of children with AD, and how these compare to those of healthy controls. In the future, it would be inter-

esting to investigate how those groups of preteens compare regarding their productive skills as well. It would be particularly interesting to explore how preteens with typical and atypical development show their confidence level in a quiz-like experiment.

5. Acknowledgments

We would like to thank the students from two schools in Eindhoven (The Netherlands), i.e., the "Pleincollege van Maerlant" and the "Pleinschool Helder", for their willingness to participate in the experiment. We thank Anniek van Doorenmalen and Lynn Verhoofstad for their help in setting up and conducting the perception experiment.

6. References

- Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Review*, 1(1), 21- 61.
- Baker, K., Montgomery, A., & Abramson, R. (2009). Perception and Lateralization of Spoken Emotion by Youths with High-Functioning Forms of Autism *Journal of Autism and Developmental Disorders* 40 (1), 123-129.
- Baron-Cohen, S. (1995) *Mindblindness: an essay on autism and theory of mind* Boston: MIT Press/Bradford books.
- Brennan, S.E. & Williams, M. (1995) The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers *Journal of Memory and Language* 34, 383-398.
- Davis, R.A.O., Bockbrader, M.A., Murphy, R.R, Hetrick, W.P. & O'Donnell, B.F. (2006) Subjective perceptual distortions and visual dysfunction in children with autism *Journal of Autism and Developmental Disorders* 36, 199-210.
- Dunn, W., Smyth-Myles, B.S. & Orr, S. (2002) Sensory processing issues associated with Asperger syndrome: A preliminary investigation *American journal of occupational therapy* 56, 97-102.
- Kanner, L. (1943) Autistic disturbances of affective contact *Nervous Child* 2, 217-250.
- Kraemer, E.J. and Swerts, M. (2006) How children and adults produce and perceive uncertainty in audiovisual speech *Language and Speech* 48, 606-639.
- Loveland K., Tunali-Kotoski B., Chen Y., Ortegon J., Pearson D., Brelsford K. & Gibbs M (1997) Emotion recognition in autism: verbal and nonverbal information. *Development and Psychopathology* 9(3), 579-93.
- Mackay, G. & Shaw, A. (2004) A comparative study of figurative language in children with autistic spectrum disorders *Child language teaching & therapy* 20, 13-32.
- Swerts, M. & Kraemer, E. (2005) Audiovisual prosody and feeling of knowing. *Journal of Memory and Language* 53, pp. 81-94.
- Visser, M., Kraemer, E. & Swerts, M. (2010) Children's expression of uncertainty in collaborative and competitive contexts *Proceedings AVSP conference*, Volterra, Italy, September 2010.
- Wilting J., Kraemer E. & Swerts M. (2006) Real vs. acted emotional speech *Proceedings Interspeech 2006*, Pittsburgh, Pa, USA, September 2006.