The relevance of visual prosody for studies in language and speech-language pathology

Swerts, M.G.J.

Published in:
International Journal of Speech-Language Pathology

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

International Journal of Speech-Language Pathology, 2009; 11(4): 282–286

informa
healthcare

COMMENTARY

The relevance of visual prosody for studies in language and speech-language pathology

MARC SWERTS

Tilburg University, The Netherlands

Abstract
I support Peppé's (2009) claim that prosody should be put on the research agenda of those working on aspects of language and speech-language pathology. But while her lead article mainly focuses on auditory forms of prosody, such as intonation, rhythm and voice quality, I argue that visual forms of prosody, in particular facial expressions, also need to be explored in this domain. Indeed, both variations in the voice and face are part of a speaker's expressive style, and are picked up as communicatively relevant cues by addressees. At the same time, there is preliminary evidence from studies of people with autism to suggest that they may have problems both with the production and comprehension of visual forms of prosody, and have difficulties to integrate input from different modalities. And finally, I propose a game-based paradigm which is potentially useful for the diagnosis and therapy of people who experience problems with the use of facial expressions in their social and linguistic interactions.

Keywords: *Visual prosody, facial expressions, autism, diagnosis, therapy.*

Introduction

The field of prosody reveals an intriguing paradox. On the one hand, most linguists would agree that prosody is an indispensable component of spoken interactions. Accordingly, there are many claims in the literature, also in the more popular media, that prosody, together with other non-verbal features (like facial expressions), accounts for a large percentage of the communication, even when it is not always entirely clear on which data such arguments are based and whether they are generally valid for all kinds of interactions (Dijkstra, Krahmer, & Swerts, 2006). The primary role of prosody would also appear from the fact that features like rhythm and intonation are acquired by young infants before they learn the words and the syntax of a language, which is often explained by findings that newborns already have access to prosody while still in the mother's womb (DeCasper & Fifer, 1980; Mehler et al., 1986). Yet, on the other hand, while it is intuitively clear that prosody has added value for human communication and is important from a developmental perspective, our knowledge of its forms and functions appears to be rather incomplete, certainly compared to what we know about other levels of linguistic structure, such as the lexicon and syntax. Typically, prosody is the component of language structure that receives comparatively little attention in educational programs.

The mismatch between the alleged importance of prosody and its ''Cinderella'' position within linguistic research (Crystal, 2009) is also apparent when looking at more specific studies in language and speech pathology. The lead article by Peppé (2009) shows how little attention prosody has so far received in this domain, despite the intuition that deficiencies in prosodic competences are likely to have negative repercussion for a person's ability to communicate with others. People who do not master the prosodic rules of a language, experience problems to express themselves in a linguistically or socially acceptable way, or may find it difficult to interpret prosodic expressions as qualifiers of another person's spoken messages. The overview paper highlights some important issues in this largely unexplored area of research, which includes a concern regarding a few methodological questions. I very much sympathize with Peppé's plea to consider both production and reception in studies of prosody and to distinguish between purely formal and functional deficiencies in prosodic competences, and I like her experimental

Correspondence: Marc Swerts, Tilburg University, Faculty of Humanities, Department of Communication and Information Sciences, Tilburg Center for Creative Computing (TiCC), PO Box 90153, NL - 5000 LE Tilburg, The Netherlands. Tel: +31 13 4662922. Fax: +31 13 4663110.
E-mail: m.g.j.swerts@uvt.nl

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Informa UK Ltd.
DOI: 10.1080/17549500902906347

framework to elicit prosodic data for diagnostic purposes.

The lead article focuses primarily on auditory forms of prosody, i.e., those aspects of non-verbal communication that can in principle be derived from the speech signal alone (like intonation, loudness, duration, and voice quality). In my reaction, I would argue that visual forms of prosody, especially facial expressions, may also be relevant for this field as they serve a range of similar communicative purposes. In the following, I will first address how such visual forms of prosody can be exploited to support specific functions. I will then embark on deficiencies in producing and perceiving facial expressions, in which I mainly focus on problems for people with autism. And before my conclusion, I will also say a few words about the diagnosis and therapy of people who experience problems with the use of facial expressions in their social and linguistic interactions.

Visual forms of prosody

Functions of visual prosody

Peppé (2009) rightly remarks that, in addition to prosody, speakers and listeners also have other linguistic devices at their disposal to support specific functions of spoken communication. Indeed, informational, attitudinal or emotional attributes of speaker utterances can be signalled by lexical variation, word order or morphemic markers as well. Along the same lines, features of ''visual prosody'' (facial expressions, hand and arm gestures, posture, or more generally variations in body language) could serve similar purposes as the auditory cues. Somewhat surprisingly, however, past studies on prosodic forms and functions have almost exclusively focused on the auditory channel alone. This is remarkable in view of the fact that in most interactions dialogue partners can both hear and see each other, so that it is only natural to expect that speakers and addressees use both voice and face for communicative purposes. There is a growing awareness that facial expressions and other forms of body language may signal communicative functions that have traditionally been attributed to variations in the speaker voice (Krahmer & Swerts, 2009).

There is of course a long tradition of research, starting with Darwin in the 19th century, into facial expressions as ''windows to the soul'', where they are viewed as correlates of a speaker's emotional state (Ekman, 2003). More recently, we are beginning to see that such expressions may signal a much wider range of communicatively relevant information. For instance, from our own work as well as that of a few others, it appears that such expressions can serve to signal the end of a sentence or a speaker turn (Barkhuysen, Krahmer, & Swerts, 2008), to highlight prominent words in an utterance (Cavé et al., 1996; Hadar et al., 1983; Swerts & Krahmer 2008; Krahmer & Swerts 2007; Dohen et al., 2004; Dohen & Lœvenbruck, 2009; Scarborough et al., 2009), to distinguish declaratives from questions (Srinivasan & Massaro, 2003), to give positive or negative feedback to an addressee (Barkhuysen, Krahmer & Swerts, 2005), to ground information in face-to-face interactions (Nakano et al., 2003), or to express a speaker's feeling-of-knowing in a question-answering situation (Krahmer & Swerts, 2005; Swerts & Krahmer, 2005).

What is more, there is evidence to suggest that the way visual features are exploited for communicative purposes is partly conventionalized. That is, speakers express themselves in ways similar to that of other people with whom they form a community. Ekman (2003), for instance, argues that cultures may differ considerably in their set of ''display rules'' that dictate which facial expressions fit certain social contexts, e.g., to show politeness or affect. Along the same lines, languages can differ in the extent to which facial cues are used for more linguistic purposes, where, for instance, it has been shown that speakers of Dutch and Italian employ facial cues differently to highlight important information in an utterance (Krahmer & Swerts, 2004). Children quickly learn such linguistic and social conventions automatically through daily practice, as they are born with a propensity to interact with others, and learn prosodic functions partly from imitating their environment (Goswami, 2008). This social perspective is evidenced by the observation that people, in their interactions with others, spontaneously adapt to each other: in the course of a dialogue, people not only start using similar words and syntactic structures as their partners (Pickering & Garrod, 2004), but also spontaneously mirror each other's facial expressions (Chartrand & Bargh, 1999). This natural tendency to align with others may constitute the basis for learning the expressive conventions of a particular community, including the use of facial expressions.

Deficiencies in visual prosody

Given the claim that a social perspective is important to learn the language- and culture-specific rules for facial expressions, one would predict that the use of such expressions is problematic for people with autism. Indeed, when someone is autistic, he or she has been claimed to be ''mind-blind'' (Baron-Cohen, 1995). As a result, people with autism have a general difficulty understanding another person's perspective, and identifying another person's thoughts and emotions. It has recently been argued that dysfunction of the mirror neuron system early in development gives rise to the impairments that characterize people with autism (Dapretto et al., 2005). This has a number of serious consequences. While they may be able to mimic another person's expressions when explicitly instructed to do so, they will not spontaneously adapt their expressions to

others, as is the case with typically developing people (McIntosh et al., 2006). Conversely, people with autism may in fact exaggerate expressions, in that they copy another person's expressions in such an extensive and atypical way (echolalia) that it does not appear to serve a clear communicative or social purpose (Rapin & Dunn, 1997). In other words, people with autism tend to have problems adapting to the expressive style of others, and therefore may have problems in ''learning'' the linguistic or social conventions for using facial expressions.

The problematic nature of facial expressions in people with autism has different forms. First, it appears that people with autism have difficulties both with the production and reception of facial expressions. As argued in the lead article, people with autism have been reported to use a monotonous or atypical kind of prosody (Baron-Cohen & Staunton, 1994; Shriberg et al., 2001). However, findings of various studies into the nature of the prosodic deficits of these people often conflict, while the receptive prosodic competences of people with autism remain largely unexplored (McCann & Peppé, 2003). With respect to the production of visual cues, gaze aversion is often cited as a characteristic of autistic children (Adrien et al., 1993; Walters et al., 1990), as well as proportionally less smiling and gesturing than typically developing children (McGee & Morrier, 2003). In reception, individuals with autism perform more poorly than others on tests of decoding facial and vocal expressions (Baron-Cohen et al., 2001; Rutherford et al., 2002), and also appear to process facial information differently from high-ability adolescents in that they are less prone to use contextual information in a face in a visual-search task (Teunisse & de Gelder, 2003).

In addition, people with autism have been reported to have great difficulty integrating information coming from the face with the auditory cues, even when results in the literature are sometimes at variance. While lip-reading skills of people with autism are comparable to those of typically developing children, there is nevertheless little influence on their speech from visual cues from the face (de Gelder et al., 1991), though others using behavioural methods have found normal audiovisual integration (Massaro & Bosseler, 2003; Williams et al., 2004). More functional studies (such as ERP studies) do show, however, that people with autism perform poorly in terms of their higher-level multisensory integration (Magnée et al., 2008a, b). So while typically developing children learn to combine input received through their ears or eyes, people with autism often keep having problems with integrating information from different sensory channels.

Diagnosis and therapy of visual prosody

The previous sections have shown that facial expressions serve different functions in human interactions, but that people with autism have difficulties with the production and interpretation of such expressions. Therefore, there is a need to develop a paradigm to find and cure problems in the use of facial expressions, using methods that yield ecologically valid data (cf. the PEPS-C test, Peppé & McCann, 2003). To this end, we are exploring to what extent games could achieve this. First, by their very nature, games represent artificial, small universes with their own rules, so that players 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 & Wehrle 1996; Salen & Zimmerman, 2003). And, since games are fun and players tend to enjoy them, there is less risk that they will induce situations which are very stressful or negative.

As an example of a game-based paradigm, let us describe a pilot experiment with typically developing younger (aged 8 years) and older (aged 12 years) children. In particular, we have constructed a card game to elicit facial expressions from participants in positive (winning) and negative (losing) contexts (Shahid, Krahmer, & Swerts, 2007). When the game starts, players see a row of six cards on a computer screen where the number of the first card is visible (''7'' in Figure 1) and the other five cards are placed upside down so the numbers are hidden. All the numbers on the cards are between 1 and 10, and a number displayed once is not repeated in a particular game. The task given to the players is to guess whether the number on the next card will be higher or lower than the previous number, and they win if they guess all cards correctly in a sequence. Unknown to the children, each game is completely deterministic, and two different alternatives are used, whereby rational choices lead to either winning or losing the game. A losing situation is shown in Figure 1 with the beginning and end state of a sequence of cards, where the final ''10'' is unexpected. The game, though very simple, turns out to be surprisingly effective (children really like it), and applicable to different age groups, different cultures, and with single and multiple players.

To investigate social aspects of audiovisual cues to emotion, the game is played either by single children or by pairs of children who sit next to each other.

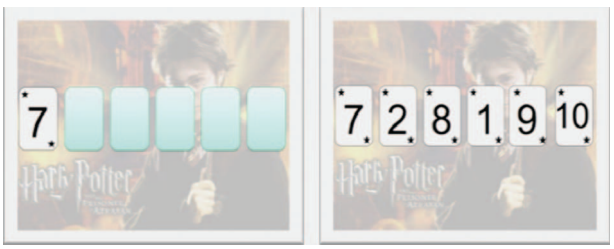

Figure 1. Illustrative stimulus materials of a card game to elicit positive and negative emotional expressions.

First analyses of recordings of children playing the game reveal that children are more expressive when they play the game together than when alone. However, it turns out that the expressive differences between single and pair conditions is bigger for older than for younger children. This could be in line with the general expectation that older children are affected more by the presence of others than younger ones. Figure 2 below shows some representative stills of a child displaying positive and negative emotions, elicited through this card game.

While we have so far conducted game-based experiments with typically developing children, it would be very interesting to try them out on children with autism as well. For a production task, one could get them to play the card game described above and video them doing it, to have a record of their facial expressions that can be compared for expressiveness with controls; one may even get them to play the game with the aim of deceiving someone as to whether they are winning/losing, to explore to what extent they are able to control their facial expressions. Additionally, in a functional receptive task, it would be interesting to invite children with autism to participate in a test in which they judge facial expressions of others as ''winning'' or ''losing''.

The above example illustrates how games can be designed such that they naturally elicit different linguistic and social contexts. In general, the proposed paradigm is in line with current interests in serious gaming (Egenfeldt-Nielsen, 2005): games are attracting increased attention as a medium for instruction and teaching, as they are considered to be rapid and cost-effective tools for simulating real-world situations. More specifically, there is a growing interest in using (computer) games in diagnosis and training programs for children with autism (Sehaba et al., 2005). The use of games potentially has a number of advantages: by playing games, people with autism may learn to recognize different contexts, to which they can adapt their facial expressions. Also, as children with autism tend to have difficulties in integrating information from multiple resources, the games can be used as a paradigm which controls for confounding factors of the environment, so that participants can focus on a specific aspect of a linguistic or social context.

Figure 2. Representative stills of a child in a winning or losing situation of the card game.

Conclusion

In this short contribution, I have tried to support Peppé's (2009) claim that prosody should be put on the research agenda of those working on aspects of language and speech-language pathology. But while her lead paper mainly focused on auditory forms of prosody, I have argued that visual forms of prosody, such as facial expressions, are very relevant as well. Indeed, both variations in the voice and face are part of a speaker's expressive style, and are jointly picked up as communicatively relevant cues by addressees. As an example of the problematic use of facial expressions, I have discussed some findings from studies of people with autism that have shown that they experience problems both with the production and comprehension of visual forms of prosody, and have difficulties in integrating input from different modalities. And finally, I have argued that we need to reflect on ecologically valid methods for the diagnosis and therapy of people who have problems with the use of facial expressions in their social and linguistic interactions.

Acknowledgement

I would like to thank Sue Peppé for her comments on a previous version of this text, and Suleman Shahid for his help in making the 2 figures.

References

Adrien, J. L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*, 617–626.

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

Barkhuysen, P., Krahmer, E., & Swerts, M. (2008). The interplay between the auditory and visual modality for end-of-utterance detection. *Journal of the Acoustical Society of America*, *123*, 354–365.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Boston, MA: MIT Press.

Baron-Cohen, S., & Staunton, R. (1994). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language*, *14*, 241–248.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The ''reading the mind in the eyes'' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 241–251.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)* (pp. 2175-2179). Philadelphia, PA.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Crystal, D. (2009). Persevering with prosody. *International Journal of Speech-Language Pathology*, *11*, 257.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Brookheimer, S. Y., & Iacoboni, M. (2005). Understanding emotions in others: mirror neuron dysfunction in chidren with autism spectrum disorders. *Nature Neuroscience*, *6*, 28–30.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*, 69–86.

DeCasper, A. J., & Fifer, W. P. (1980). On human bonding: Newborns prefer their mother's voices. *Science*, *208*, 1174–1176.

Dijkstra C., Krahmer E., & Swerts, M. (2006). Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. *Proceedings of the Speech Prosody 2006 Conference.* Dresden, May.

Dohen, M., & Lœvenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, *52*, 177–206.

Dohen, M., Lœvenbruck, H., Cathiard, M.-A., & Schwartz, J.-L. (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, *44*, 155–172.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment. Exploring the educational potential of computer games.* PhD thesis, University of Copenhagen.

Ekman, P. (2003). *Emotions revealed. Understanding faces and feelings.* London: Weidenfeld & Nicolson.

Goswami, U. (2008). *Cognitive development. The learning brain.* Hove: Psychology press.

Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech*, *26*, 117–129.

Kaiser, S., & Wehrle, T. (1996). Situated emotional problem solving in interactive computer games. In N. H. Frijda (Ed.), *Proceedings of the IXth Conference of the International Society for Research on Emotions* (pp. 276–280). Toronto.

Krahmer, E., & Swerts, M. (Eds.) (2009). Special double issue on Audiovisual Prosody. *Language and Speech*, *52*, 129–386.

Krahmer, E., & Swerts, M. (2004). More about brows: a cross-linguistic study via analysis-by-synthesis. In Z. Ruttkay, et al. (Eds.), *From brows to trust: Evaluating embodied conversational agents* (pp. 191–216). Dordrecht: Kluwer.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*, 29–53.

Krahmer, E., & Swerts, M. (2007). The effect of visual beats on prosodic prominence: acoustic analyses, auditory perception, and visual perception. *Journal of Memory and Language*, *57*, 396–414.

Magnée, M. J. C. M., de Gelder, B., van Engeland, H., & Kemner, C. (2008a). Atypical processing of fearful face-voice pairs in pervasive developmental disorder: An ERP study. *Clinical Neurophysiology*, *119*, 2004–2010.

Magnée, M. J. C. M., de Gelder, B., van Engeland, H., & Kemner, C. (2008b). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *The Journal of Child Psychology and Psychiatry*, *49*, 995–1000.

Massaro, D. W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal on Developmental and Learning Disorders*, *7*, 111–144.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, *38*, 325–350.

McGee, G., & Morrier, M. (2003). Clinical implications of research in nonverbal behaviour of children with autism. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *Nonverbal behaviour in clinical settings.* Oxford: Oxford University Press.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science*, *9*, 295–302.

Mehler, J., Lambertz, G., Jusczyk, P. W., & Amiel-Tison, C. (1986). Discrimination de la langue maternelle par le nouveau-né. *Comptes rendus de l'Académie des Sciences de Paris 303, série, III*, 637–640.

Nakano, Y. I., Reinstein, G., Stocky, T., & Cassell, J. (2003). Towards a model of face-to-face grounding. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Sapporo, Japan.

Peppé, S. J. E. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology*, *11*, 258–271.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics*, *17*, 345–354.

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.

Rapin, I., & Dunn, M. (1997). Language disorders in children with autism. *Seminars in Pediatric Neurology*, *4*, 86–92.

Rutherford, M.D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: Sensitivity in childhood psychopathology. *Journal of Autism and Developmental Disorders*, *32*, 189–194.

Salen, K., & Zimmerman, E. (2003). *Rules of play. Game design fundamentals.* Cambridge, MA: MIT Press.

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., Alwan, A., & Auer, E. T. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, *52*, 135–175.

Sehaba, K., Estraillier, P., & Lambert, D. (2005). Interactive educational games for autistic children with agent-based system. *Lecture Notes in Computer Science (LNCS)*, *3711*, 422–432.

Shahid, S., Krahmer, E., & Swerts, M. (2007). Audiovisual emotional speech of game playing children: Effects of age and culture. *Proceedings of Interspeech 2007.* Antwerp, August.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger's Syndrome. *Journal of Speech, Language, and Hearing Research*, *44*, 1097–1115.

Srinivasan, R. & Massaro, D. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, *46*, 1–22.

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

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*, 219–238.

Teunisse, J.-P., & de Gelder, B. (2003). Face processing in adolescents with autistic disorder: The inversion and composite effects. *Brain and Cognition*, *52*, 285–294.

Walters, A. S., Barrett, R. P., & Feinstein, C. (1990). Social relatedness and autism: Current research, issues, directions. *Research in Developmental Disabilities*, *11*, 303–326.

Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities*, *25*, 595–575.