Autistic individuals benefit from gestures during degraded speech comprehension

Running title:
Gestural enhancement of speech in autism

Authors and affiliations:
The Communicative Brain (Sara Mazzini[1*], Noor Seijdel[1*] & Linda Drijvers[1,2*])

*All authors contributed equally to this work

[1] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

Corresponding author:
Sara Mazzini
Neurobiology of Language Department - The Communicative Brain
Max Planck Institute for Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
E-mail: sara.mazzini@mpi.nl

Acknowledgements
We thank the Neurobiology of Language department for helpful discussions. This work was funded by a Minerva Fast Track Fellowship from the Max Planck Gesellschaft to LD.

Lay Summary

Our study explored how gestures, alongside spoken words, can help people with autism understand speech, especially when the speech quality is poor. We found that both autistic and non-autistic individuals use meaningful information from hand gestures when understanding unclear speech.

Keywords

speech-gesture integration, autism, degraded speech comprehension, multimodal language

Abstract

Iconic co-speech gestures enhance degraded speech comprehension in neurotypical adults. Nonetheless, the benefit of gestures in comprehending degraded speech has not been investigated in neurodivergent populations, such as autistic individuals. Previous research demonstrated atypical audiovisual and speech-gesture integration in autistic individuals, suggesting that integrating speech and gestures may be more challenging and less beneficial for speech comprehension in adverse listening conditions in comparison to neurotypicals. Conversely, autistic individuals could also benefit from additional cues to comprehend speech in noise, as they encounter difficulties in filtering relevant information from noise. In the present study, we investigated gestural enhancement of degraded speech in neurotypical and autistic adults. Participants were presented with videos of an actress uttering a Dutch action verb and had to complete a 4-alternative forced choice task. The action verb was produced in either clear or degraded speech and accompanied by a matching gesture or without a gesture. We observed a gestural enhancement effect in both neurotypical and autistic individuals, and no difference in the size of this effect between the groups. Our findings suggest that despite the previously reported differences in audiovisual integration and gesture interpretation, autistic individuals do benefit from gestures in degraded speech comprehension, similarly to neurotypicals. These findings provide relevant insights to improve communication practices with autistic individuals and to develop new interventions for speech comprehension.

Introduction

Human communication involves more than auditory speech comprehension: interlocutors must also pay attention to and integrate communicative signals from visual bodily signals, such as hand gestures. Specifically iconic gestures, which can be described as hand gestures that illustrate object attributes, actions, and space (McNeill, 1992), have been demonstrated to improve speech comprehension in adverse listening conditions (Drijvers & Özyürek, 2017). This gestural enhancement effect has been replicated in multiple studies including various neurotypical populations (e.g., Drijvers et al., 2019, 2021; Schubotz et al., 2021). Nonetheless, the benefit of gestures in degraded speech comprehension has not yet been investigated in neurodivergent populations, which commonly encounter socio-communicative difficulties, such as in individuals with autism spectrum disorder (ASD). We here fill this gap in the literature.

Communicative difficulties have often been observed in autistic children, varying on a continuum of severity: these may include a limited ability to understand and produce speech, as well as to understand and produce non-verbal information (e.g., Vogindroukas et al., 2022). Unisensory and multisensory impairments have also been observed in autistic children (Siemann et al., 2020), but results on this have been mixed. For example, unisensory processing in ASD has been reported to be both enhanced (e.g., visually: Bertone et al., 2005; auditory: Bonnel et al., 2003), and reduced (e.g., Blake et al., 2003), depending on the complexity of the stimuli (Bertone et al., 2005). In both unisensory and multisensory contexts, low-level stimuli were demonstrated to be processed relatively efficiently, but more complex stimuli such as speech were not (Mottron et al., 2006; Stevenson et al., 2014).

In multimodal contexts, multisensory integration of both speech (Foxe et al., 2015; Smith & Bennetto, 2007) and non-speech stimuli (i.e., flashes and tones, see Beker et al., 2018)

has been demonstrated to be impaired in autistic children, although both seemed to improve throughout adolescence (Beker et al. 2018; Foxe et al. 2015). For example, adolescents with ASD showed an attenuated enhancement effect of visual speech during speech comprehension (Smith & Bennetto, 2007), that improved as they aged (Foxe et al., 2015).

To our knowledge, only a few studies have investigated multisensory integration in ASD using speech and hand gestures: Trujillo et al., (2021) reported that autistic individuals were overall unimpaired when recognizing gestures, but they interpret more complex movements differently than neurotypical adults do. Moreover, Silverman et al., (2010) demonstrated that autistic children and adolescents were slower to respond to audiovisual stimuli than to audio-only stimuli, which suggests that hand gestures may hinder, and not benefit, speech comprehension. It is currently unclear whether this is overcome throughout adolescence, and whether or not this differs for situations where the semantic information that is conveyed by iconic gestures could improve speech comprehension, such as when speech is embedded in adverse listening conditions.

In the present study, we directly tested gestural enhancement of degraded speech in neurotypical and autistic adults, to better understand how autistic adults make use of gestural information and how this visual semantic information facilitates speech recognition in adverse listening conditions. In line with previous literature, we expected a gestural enhancement effect for neurotypical adults (Drijvers & Ozyurek, 2017). For individuals with ASD, we expected a smaller gestural enhancement effect, based on the difficulties that have been observed in this group with processing both auditory (Ronconi et al., 2023) and visual (e.g., Silverman, 2010; Trujillo et al., 2021) information. However, some work suggested that these multisensory difficulties are overcome in adolescence (Foxe et al., 2015; Beker et al., 2018). We therefore expected some gestural enhancement of degraded speech comprehension to occur, albeit smaller than the effect we hypothesized for neurotypical adults.

Methods

Participants

160 individuals (116 females, 18-40 years old, all native speakers of Dutch) participated in an online experiment. All participants were recruited from the participant database of the Max Planck Institute for Psycholinguistics, as well as through social media and forum advertisements. Data collection was part of a larger research line on audiovisual communication, and as part of that project, participants were assigned to one of four groups: 1) neurotypical controls, 2) ADHD, 3) dyslexia or 4) autism. For the present study, 40 healthy neurotypical controls (30 females, mean age = 24.1, SD = 4.17) and 40 autistic participants (26 females, mean age = 26.83, SD = 6.02) were included, resulting in 80 participants who were included in the analyses. All participants reported no hearing impairments and normal or corrected-to-normal vision. All participants gave written informed consent and received monetary compensation for participation.

Stimuli

The videos used in the experiment were selected from a larger pool of 240 videos previously employed in studies by Drijvers and Ozyürek (2017), Drijvers, Ozyürek, and Jensen (2018), and Drijvers, Jensen, and Spaak (2021). Participants were presented with 160 video clips of an actress who uttered an action verb accompanied by an iconic gesture or no gesture (see for a detailed description of pretests on recognizability and iconicity of the gestures, Drijvers & Ozyürek, 2017), while the speech in the video was presented clearly or with 6-band noise-vocoding (see Drijvers & Ozyurek, 2017). All gestures used in the videos were rated as potentially ambiguous when viewed without speech, indicating that they were not pantomimes (see for rationale: Drijvers & Ozyurek, 2017).

Experimental procedure

The experiment was conducted online using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Participants were instructed to wear headphones. The experiment began with a sound check, to ensure a comfortable volume level and to verify that the participants were using headphones. Following these checks, participants completed questionnaires on demographics. In the main experiment, participants performed a free-recall task. Each trial commenced with a fixation cross (1000 ms), followed by the video (2000s ms). We asked participants to press the space bar as quickly as possible when they recognized the verb that was being conveyed by the actress in the video. After this, the participants could freely type the verb they thought was being conveyed. All videos were presented in a randomized order and divided over blocks of 40. Participants could take a self-paced rest after every block. All participants completed the experiment within 45 minutes.

Analysis

For both groups, we calculated the Token Sort Ratio (TSR) for all responses, to get a more detailed picture of how similar a participant's response was to the correct answer. TSR uses fuzzy string matching to calculate the similarity between two strings (e.g., the correct answer, and a participant's answer), and returns a percentage overlap between 0 and 100. All responses that were slower than 4000 ms and faster than 300 ms (7,2% of the data, 920 entries) were excluded. These responses included question marks and short statements that indicated that the participant did not understand the speech in the video.

We then tested whether a gestural enhancement effect of degraded speech comprehension existed for both groups. To this end, we fitted linear mixed effects models (using the lme4 package, Bates et al., 2015) in R (version 4.2.3, R Core Team, 2023). P-values were obtained using the lmerTest package (v3.1.1., Kuznetsova et al., 2017). All models included a maximal random effects structure. When the model did not converge, we used the

estimates of this model fit as starting values to restart the fit, and compared the estimates from different optimizers using the allFit() function. If these yielded similar values, warnings about singularity were considered false positives. This was the case for all model fits. To investigate whether the two groups differed in their gestural enhancement effect, we then ran another model including a three-way interaction between Speech, Gesture and Group, again using a maximal random effects structure.

Results

Gestures enhance degraded speech comprehension for both neurotypical and autistic participants

Both neurotypical and autistic participants performed at ceiling in the clear speech condition, independent of gesture presence (No gesture: mean TSR = 99.00% (NT) / 99.03% (A), SD = 6.39% (NT) / 6.16% (A); Gesture: mean TSR = 97.83% (NT) / 97.88%, SD = 11.40% (NT), 10.87% (A)). When speech was degraded and gestures were present in the video, both neurotypical and autistic participants performed better than when speech was degraded and no gesture was present (No gesture: mean TSR = 64.17% (NT) / 60.27% (A), SD = 33.65% (NT) / 36.27% (A); Gesture: mean TSR = 80.63% (NT) / 74.90%, SD = 29.95% (NT), 34.97% (A)). Both groups demonstrated a gestural enhancement effect of degraded speech comprehension (Neurotypicals: β = 17.38, *SE* = 4.06, t(105.25)= 4.26, p = <.001; Autistic participants: β = 15.04, *SE* = 4.65, t(100.48)= 3.13, p = .001). The size of this effect did not differ between the two groups (β = 0.58, *SE* = 42.29, t(11640.49) = 0.11, p = .99, see Figure 1).

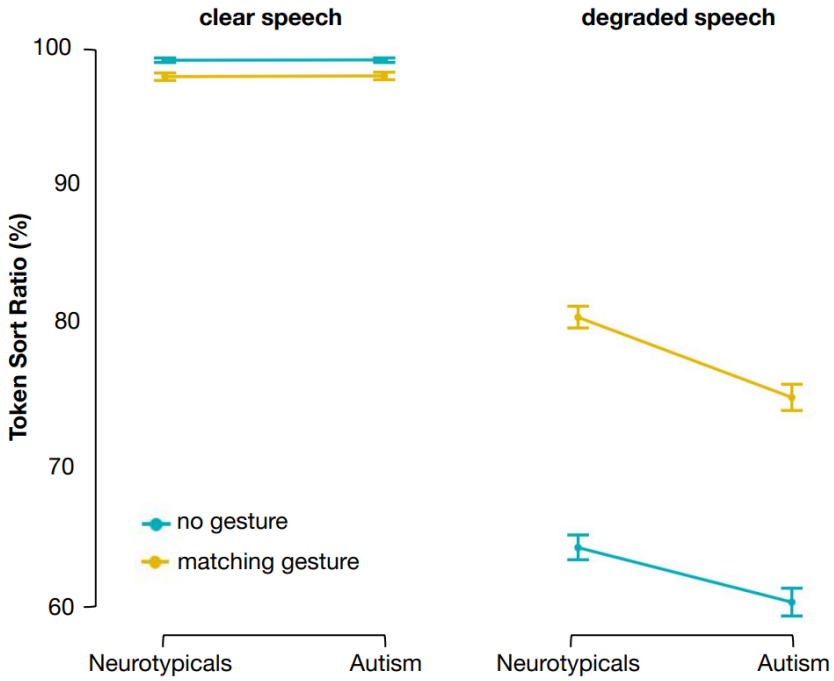

Figure 1: Gestures enhance degraded speech comprehension for both the neurotypical and autism group. Token Sort Ratio (%) per condition. Error bars represent SE.

Discussion

In the current study, we asked whether a gestural enhancement effect of degraded speech comprehension exists in autistic individuals and whether this effect differed from the gestural enhancement effect that has been typically observed in neurotypical adults (Drijvers et al., 2019; Drijvers & Özyürek, 2017; Schubotz et al., 2021; Wilms et al., 2022). Our findings provide evidence that both groups benefited from the presence of gestures when speech was degraded and that this effect did not differ between groups. This suggests that iconic gestures can facilitate speech recognition in adverse listening conditions for individuals with ASD, similar to what has been observed for neurotypical individuals. These results have important implications for understanding the role of multimodal signals in speech comprehension for individuals with ASD. We discuss these results in more detail below.

We extended findings from previous work in neurotypical populations (Drijvers & Ozyurek, 2017; Schubotz et al., 2015; Drijvers, Van der Plas, Jensen, 2019; Wilms et al., 2022) by demonstrating that individuals with ASD show a significant enhancement in degraded speech comprehension when accompanied by iconic gestures. Contrary to our hypothesis that this effect may be reduced for autistic individuals due to their known difficulties in processing non-verbal information (e.g., Silverman et al., 2010), we found that this effect did not differ from neurotypical adults.

These results also do not concur with results from Silverman et al., (2010), who observed impaired gesture-speech integration in autistic individuals. Nonetheless, this divergence in results may be explained by differences in the experiment design: first, our sample included adult participants, whereas in Silverman et al. (2010) children and adolescents were included. As previous findings on visual speech (e.g., Foxe et al., 2015) suggested that multisensory integration difficulties in individuals ameliorate during adolescence and adulthood, the gestural enhancement effect that we have observed in our study may have developed with age. Secondly, the gestures used by Silverman (2010) were redundant to the speech presented, whereas in our study gestures provided complementary information to speech. We therefore stipulate that a gestural enhancement effect in autistic individuals may be especially prevalent when there is a necessity to integrate stimuli coming from different modalities to understand speech, such as in adverse listening conditions.

In the current work, we can exclude neither of these two options. Overall, our results extend what was previously reported by Silverman et al (2010) to an older population and propose an interesting pattern, which should be tested and confirmed by future research: the presence of gestures facilitates speech comprehension in autistic individuals, whenever these gestures provide complementary information to speech, and this effect may develop with age.

Limitations and future directions

Several limitations of the present study should be acknowledged. Firstly, as discussed above, the current work only focused on autistic individuals of adult age. As previous work has suggested that for individuals with autism developmental differences may exist in multisensory integration (Beker et al., 2018; Foxe et al., 2015; Silverman et al., 2010), which ameliorate by early adulthood, future work could consider including different age groups to investigate whether similar effects occur for more complex visual information, such as the iconic gestures used in the current study. Second, the current work did not differentiate between different parts of the autistic spectrum, limiting the generalizability to all autistic adults. Third, we did not collect any data on the potential communicative difficulties that autistic individuals may experience during daily life. For example, autistic individuals may have found it more difficult to filter relevant information from noise (Ronconi et al., 2023; Schelinski & von Kriegstein, 2020). Future work could investigate this by, for example, including questionnaires targeted at processing sound and speech in adverse listening conditions, or by including the actions and feelings questionnaire, which measures how well someone uses and understands visual communicative signals (van der Meer et al., 2022).

Conclusion

We here demonstrate that autistic individuals benefit from visual semantic information conveyed by gestures during degraded speech comprehension. This enhancement effect is not different from the enhancement effect that neurotypical adults experience. These results therefore highlight the benefit of using visual bodily signals during communication in natural, adverse listening conditions, and invite possibilities of incorporating this knowledge in interventions targeting (degraded) speech comprehension in autistic adults.

Data availability statement

All data and code is available at https://osf.io/7fgjt/

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience & Biobehavioral Reviews*, *84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain: A Journal of Neurology*, *128*(Pt 10), 2430–2441. https://doi.org/10.1093/brain/awh561

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, *14*(2), 151–157. https://doi.org/10.1111/1467-9280.01434

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Drijvers, L., Jensen, O., & Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping*, *42*(4), 1138–1152. https://doi.org/10.1002/hbm.25282

Drijvers, L., & Özyürek, A. (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Drijvers, L., van der Plas, M., Özyürek, A., & Jensen, O. (2019). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *NeuroImage*, *194*, 55–67. https://doi.org/10.1016/j.neuroimage.2019.03.032

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H.-P., Russo, N. N., Blanco, D., Saint-Amour, D., & Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(2), 298–312. https://doi.org/10.1093/cercor/bht213

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago press.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Ronconi, L., Vitale, A., Federici, A., Mazzoni, N., Battaglini, L., Molteni, M., & Casartelli, L. (2023). Neural dynamics driving audio-visual integration in autism. *Cerebral Cortex (New York, N.Y.: 1991)*, *33*(3), 543–556. https://doi.org/10.1093/cercor/bhac083

Schelinski, S., & von Kriegstein, K. (2020). Brief Report: Speech-in-Noise Recognition and the Relation to Vocal Pitch Perception in Adults with Autism Spectrum Disorder and

Typical Development. *Journal of Autism and Developmental Disorders*, *50*(1), 356–363. https://doi.org/10.1007/s10803-019-04244-1

Schubotz, L., Holler, J., Drijvers, L., & Özyürek, A. (2021). Aging and working memory modulate the ability to benefit from visible speech and iconic gestures during speech-in-noise comprehension. *Psychological Research*, *85*(5), 1997–2011. https://doi.org/10.1007/s00426-020-01363-8

Siemann, J. K., Veenstra-VanderWeele, J., & Wallace, M. T. (2020). Approaches to Understanding Multisensory Dysfunction in Autism Spectrum Disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *13*(9), 1430–1449. https://doi.org/10.1002/aur.2375

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–393. https://doi.org/10.1016/j.cognition.2010.01.002

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(8), 813–821. https://doi.org/10.1111/j.1469-7610.2007.01766.x

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Trujillo, J. P., Özyürek, A., Kan, C. C., Sheftel-Simanova, I., & Bekkering, H. (2021). Differences in the production and perception of communicative kinematics in autism. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(12), 2640–2653. https://doi.org/10.1002/aur.2611

van der Meer, H. A., Sheftel-Simanova, I., Kan, C. C., & Trujillo, J. P. (2022). Translation, Cross-Cultural Adaptation, and Validation of a Dutch Version of the Actions and Feelings Questionnaire in Autistic and Neurotypical Adults. *Journal of Autism and Developmental Disorders*, *52*(4), 1771–1777. https://doi.org/10.1007/s10803-021-05082-w

Vogindroukas, I., Stankova, M., Chelas, E.-N., & Proedrou, A. (2022). Language and Speech Characteristics in Autism. *Neuropsychiatric Disease and Treatment*, *18*, 2367–2377. https://doi.org/10.2147/NDT.S331987

Wilms, V., Drijvers, L., & Brouwer, S. (2022). The Effects of Iconic Gestures and Babble Language on Word Intelligibility in Sentence Context. *Journal of Speech, Language, and Hearing Research: JSLHR*, *65*(5), 1822–1838. https://doi.org/10.1044/2022_JSLHR-21-00387