Effects of Eye Gaze Direction on Vocal Imitation in Individuals with Autism Spectrum Disorder: Does Joint Attention Matter?

Marie Postma-Nilsenová (M.Nilsenova@tilburguniversity.edu)

Mariska van Kastel (mvkastel7@gmail.com)

Martijn Balsters (M.J.H.Balsters@tilburguniversity.edu)

Tilburg center for Cognition and Communication (TiCC), Tilburg University, The Netherlands

Abstract

Several recent studies showed the effect of eye gaze direction on both instructed and spontaneous imitative behavior, as well as the acquisition of action-effect binding. In particular, direct eye gaze of a model gesturer/talker, compared to averted eye gaze, gives rise to faster gesture imitation and better vocal imitation, and reinforces intersubjective stimulus-effect learning. In an experiment with Autism Spectrum Disorder participants and a control group (N = 32), we explored vocal imitation in conditions with engaged eye gaze, averted gaze and gaze establishing joint attention. We found that speakers from both groups were least likely to mimic the vocal patterns of the model talker in a condition with joint attention. The finding suggests that establishing joint attention by gaze directing negatively affects vocal imitative behavior.

Keywords: Eye gaze; Vocal Imitation; Mimicry; Joint Attention; Autism.

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, characterized by impairments in social and communicative behavior, and accompanying restricted range of interests and behaviors, that affects approximately 1% of the population (Tesink, 2013). Despite the great varieties in symptom severity and levels of intelligence, delayed information processing and difficulties regarding social interactions appear to be common in all individuals diagnosed with the disorder. With regard to interaction with other people, ASD individuals are known to have trouble with engaging in social interactions, responding to external emotional cues and making eye contact with others (Werner et al., 2000).

Arguably, one of the most important skills regarding social interaction and communication of affect concerns processing of nonverbal cues and their imitation. At a very early stage of life, social behaviors such as engaging in eye contact, joint attention and facial as well as vocal expression recognition are present in normally developing young infants. Nonverbal cues provide infants with important information, helping them with communication resulting in increased and better survival. Infants are exposed to intonation patterns even before they are born and display facial recognition and a visual preference for faces already during the first six months of life (Dawson et al., 2005). During their first year, they become capable of interpreting features such as gaze direction, facial movements and vocal expressions of emotion. Given the fact that the aforementioned behaviors have been claimed to be impaired in people with autism, these deficiencies might play an important diagnostic role in terms of early indication of brain functioning typical for autism and in possible communication training.

Interestingly, several studies have shown that, compared to normal individuals, autistic individuals use different strategies to process faces (Dawson et al., 2005; Langdell, 1978). Whereas normally developed individuals use a more holistic approach in terms of facial expression recognition, autistic individuals seem to rely on a more feature-based piecemeal information processing. For instance, when performing facial identification tasks, autistic children have shown a preference for mouths (lower facial parts) rather than the eyes (higher facial parts) compared to typically developed individuals (Langdell, 1978).

Next to that, ASD individuals perform worse at brief stimulus presentations, suggesting a slower processing of gaze direction compared to non-autistic individuals (Wallace et al., 2006). Past eye-tracking studies also showed that autistic individuals fixate less frequently and for a shorter time on others’ eyes than non-ASD individuals (Senju, 2013; Speer et al., 2007). In several experiments, autistic individuals failed to show the ‘eye contact effect’, a phenomenon in which perceived gaze is immediately followed by the processing of social information (Senju, Hasegawa, & Tojo, 2005; Senju, & Johnsson, 2009).

Nevertheless, although atypical gaze behavior is one diagnostic feature of ASD, previous experiments suggest that autistic individuals do not suffer from a general impairment in gaze processing; for instance, they can tell direct from averted gaze and are able to discriminate subtle differences in gaze direction (Leekam et al., 1997; Senju, 2013; Baron-Cohen et al., 1995). In addition, ASD individuals also demonstrate
the so-called gaze cueing effect, the phenomenon in which the observer shifts his attention to the direction of the perceived gaze (Frischen, Bayliss, & Tipper, 2007; Senju et al., 2004). Other investigations showed that ASD individuals are equally fast at detecting faces and at gender recognition, regardless of gaze direction (von Grünau, & Anston, 1995; Pellicano, & Macrae, 2009). In addition, physiological studies demonstrated enhanced responses in autistic individuals to direct perceived gaze. Strikingly, children with ASD showed a stronger skin conductance response for straight gaze compared to averted gaze, whereas typically developed children showed no differences in both conditions (Kylliäinen, & Hietanen, 2006) which would suggest that they are more sensitive to eye contact.

The studies summarized above suggest that although gaze processing is not fundamentally impaired in ASD and perceived direct gaze can evoke physiological reactions in autistic individuals, it does not spontaneously facilitate cognitive or behavioral processing (Senju, 2013). In line with this generalization, ASD individuals are also known to perform worse with respect to imitative behavior, compared to non-ASD individuals. As indicated by experimental results obtained in different domains and with varying methods, there appears to be a strong relation between eye gaze and imitation. Wang, Newport, and Hamilton (2011) found that reaction times for displaying gestures congruent with those shown on a video were faster if the person in the video engaged the participant with her gaze compared to videos where she was looking away (replication of the result reported in Wang, Ramsey, & Hamilton, 2011). Faster reaction times were also reported by Sato and Itakura (2013) for intersubjective action-effect binding acquired in a condition with target’s direct gaze, compared to averted gaze or closed eyes. A study conducted by Postma-Nilsenová, Bruninkhuis, and Postma (2013), combining both eye contact and vocal mimicry in typically functioning participants, revealed a sensitivity to the eye gaze direction of an embodied interaction agent (EIA); participants adjusted their pitch slope more to the EIA in case of direct eye gaze than with averted gaze.

Imitation (also sometimes referred to as adaptive behavior, or accommodation, though not to be confused with mimicry, see Gonzalez-Lienres, Shamy-Tsoory, & Brüne, 2013) stands for the unconscious tendency to observe other one’s behavior and copy it while being aware of the difference between self and other (Blake more & Frith, 2005; Chartrand & Van Baaren, 2009; Gonzalez-Lienres et al., 2013; Van Holland, 2007). According to previous research, imitation of the interlocutor is one of the most important aspects in social communication (Ingersoll, 2008; Toth, Munson, Meltzoff & Dawson, 2006). It serves several general functions, such as providing a sense of mutual connectedness, shared social experiences and a means of communication between social partners (Meltzoff, 2005; Toth et al., 2006). Additionally, an interlocutor is evaluated more positively in a conversation when he or she imitates the interaction partner (Chartrand & Barg, 1999). In short, imitation is a powerful mechanism that both tests and strengthens social relationships; besides, it is closely related to cooperative behavior and is often seen as a form of empathy (Bensing, 1991).

In typically developing children, the ability to imitate is present early on in life. It is a learning strategy that children use to acquire and control new behavior and whereby the goal of the behavior of the interaction partner is understood (Van Holland, 2007). Children with a form of autism, however, have difficulty with understanding the goal of their interaction partners. Although they may mimic the same behavior as typically developing children in an interaction, they do not understand the underlying social meaning of that particular behavior (Van Holland, 2007; Williams et al., 2006). Also as adults, autistic individuals perform worse than typically functioning individuals with respect to some types of imitative behavior (Hadjikhani, 2007; Rogers, Hepburn, Stackhouse & Wehner, 2003).

The complex empirical results regarding imitative behavior displayed by autistic individuals can be linked to different theoretical notions: 'functional adaptation' and 'automatic imitation': Functional adaptation serves to support audience design and common ground during a conversation and it is usually employed consciously. Automatic imitation, on the other hand, is mostly an unconscious process. The communication deficits described for people with autism commonly concern functional adaptation, to be precise audience design accommodation (Slocombe et al., 2012). This is in contrast to automatic imitation; while autistic individuals have difficulty with understanding the interaction partner’s behavior goal, they can and do imitate that behavior (Van Holland, 2007; Williams et al., 2006), e.g., facial expressions (Press, Richardson & Bird, 2010). A.o., Slocombe et al. (2012) investigated in their study the linguistic imitation in adults with and without Asperger’s syndrome. Their results revealed that adults with Asperger’s syndrome demonstrated linguistic imitation in terms of lexical choice, spatial frame of reference and syntactic structure. In addition, it appeared that there was no difference between adults with and without Asperger’s syndrome.

Despite the vast literature on autism and joint attention on the one hand, and autism and imitation on the other hand, to our knowledge, no studies have been conducted so far in which effects of eye gaze on vocal imitation were investigated among individuals with ASD. In the present study, we explored imitative vocal behavior in three conditions, with the target displaying a direct (engaging) eye gaze, with the target averting her eyes without establishing eye contact, and, finally, with her initiating the participant’s gaze following by first establishing eye contact and then directing her gaze towards a relevant object in the environment. In view of earlier findings, we examined if (1) adult ASD partici-
pants would imitate the prosody of the target’s voice to a lesser degree than typically functioning participants, and (2) if the degree of imitation will be affected by the target’s gaze behavior, particularly in a comparison between the typically functioning group and the ASD group.

**Experiment**

The goal of the experiment was to investigate if the direction of model talker’s eye gaze affects vocal pitch imitation in speakers with ASD.

**Method**

**Participants** Thirty-two participants (28 male, 4 female, divided equally per group), half of them diagnosed with autism and the other half representing a control group, were recruited for the study. The control group matched the ASD group on age, education and gender (for ASD: Mean Age = 37.9, SD = 13.4, age range 21-67; for Control: Mean Age = 36.6, SD = 15.4, age range 18-69). The uneven number of male and female participants is due to the fact that autism is more common in males than in females (Tesink, 2013). The ASD participants were highly functioning individuals who were previously diagnosed by a medical specialist and were clients of two different institutions in the Netherlands; they all scored above average on the Autism Spectrum Quotient form (Baron-Cohen, 2012), as reported by the institutions through which the participants were recruited (scores were not made available to us). This study received ethical approval from the Department of Psychology, University Ethical Review Board.

**Design** The experiment had a 2 x 3 within-between participant design, with ASD and typically functioning participants and with the variable Eye Gaze consisting of three experimental conditions: Direct Gaze, Joint Attention (initial direct gaze directed towards an object in the left or right visual field) and Averted Gaze (gaze averted left or right without any additional object without an initial direct gaze), see Figure 1. The object in the Joint Attention condition was the digit pronounced by the model talker. The dependent variable was the degree of pitch imitation, measured for Pitch Slope as well as imitation with respect to the duration of voiced frames, a measure of Speech Rate.

**Material** The experimental session consisted of 60 randomized trials. The participant first observed an audiovisual clip in which the model talker pronounced a monosyllabic digit (ten different digits in total) and subsequently repeated the digit. In the beginning of the session, the participant’s vocal baseline was recorded; the input for the baseline was presented visually on the screen.

The audiovisual clips showed a female model talker filmed against a neutral background. The vocal stimuli were recorded with five different intonation patterns: rise, late rise, fall, low and high (see Figure 2). All clips had the duration of 10 seconds, including a 0.5 sec fade-from-black and fade-to-black to smooth the transition between consecutive movies, to mark the beginning and end of each stimulus and to avoid unnatural gaze duration. The voicing lasted approximately 400 ms followed by a response window during which the participants were instructed to repeat the digit uttered by the model talker. In the Joint Attention condition, the digit appeared on the screen after initial eye contact, followed by a head turn of the model talker and by her uttering the digit. Prior to the experiment, the recordings were tested for being emotionally neutral; recordings that were seen as emotionally expressive (1 visual and 3 audio) were removed from the stimulus set and re-recorded. In the Joint Attention condition, the digits appeared in the left or right visual field after three seconds and before the model talker pronounced the digit.

**Procedure** The experiment was conducted with the help of the E-prime environment (Psychological Software Tools Inc., Pittsburgh, Pennsylvania, USA) on a Dell E173pc computer with no Internet connection and virus detection to prevent automatic pop-ups and updates and with a PC 320 G4ME headset. At the start of the experiment, the participants received a

---

1 In total, data were collected from 24 ASD participants; only the results of the participants for which a matched control participant was available are reported here. A splitfile repeated measures analysis of variance with the ASD group revealed exactly the same patterns for the two dependant variables as those reported in the results section, with a comparable level of significance.
Each trial was introduced by a 500 ms fixation cross. The 60 trials were divided into two blocks with a two-minute break in between. The recordings of the participant's vocal output were stored in separate uncompressed wav-files. They were analyzed with the help of the Praat speech analysis software (Boersma & Weenink, 2014) using Praat scripts with standardized pitch tracking setting for male (60-300 Hz) and female (80-500 Hz) speakers. To detect if participants changed their vocal parameters differently in the three experimental conditions, we calculated for each digit the absolute difference between the measurement in the baseline condition and the measurement in the condition with Joint Attention, Direct Gaze and Averted Gaze, using the vocal parameters Pitch Slope (measuring local contour changes) and Speech Rate (measured by counting voiced frames in the recording based on the detectable pitch). In order to calculate the degree of imitation, each measurement was compared to the model talker’s parameters, in the following way: First, we calculated the absolute difference between the model talker’s Pitch Slope/Speech Rate and the participant’s baseline, i.e., the Slope/Rate of the participant before being exposed to the model talker. Second, the absolute difference between the model talker’s Slope/Rate and the participant’s Slope/Rate while repeating after the model talker was calculated. Finally, by subtracting the values of the baseline difference and the repetition difference, we obtained a measure of vocal imitation. For each participant, we calculated the mean degree of imitation in the three experimental conditions.

With respect to imitation of Speech Rate, there was no main effect of experimental group (ASD, Control) on Speech Rate, $F(1.30) = 0.05$, $p = .826$ and no interaction effect between experimental group and Gaze Direction (Direct Gaze, Averted Gaze, Joint Attention), $F(2,60) = 0.46$, $p = .636$. There was a significant main effect of Gaze Direction, $F(2,60) = 3.35$, $p = .042$, partial eta squared $= .10$. Participants imitated the most in the condition with Averted Gaze and the least in the condition with Joint Attention. Pairwise comparisons with Bonferroni correction revealed a significant difference between imitation in the condition with Direct Gaze and with Joint Attention, $t(37) = 1.97$, $p = .056$; there was no difference between Direct and Averted Gaze, $t(37) = 0.28$, $p = .783$, and between Averted Gaze and Joint Attention, $t(37) = 1.78$, $p = .083$. For descriptives, see Table 1.

Table 1: Descriptive values for Pitch Slope Imitation (measured in Hz) in the three experimental conditions (positive value indicates convergence to target’s values, negative value shows divergence).

<table>
<thead>
<tr>
<th>Group</th>
<th>EMM</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct Gaze</td>
<td>22.5</td>
<td>27.2</td>
</tr>
<tr>
<td>Averted Gaze</td>
<td>10.1</td>
<td>26.7</td>
</tr>
<tr>
<td>Joint Attention</td>
<td>-29.3</td>
<td>29.1</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct Gaze</td>
<td>21.3</td>
<td>27.2</td>
</tr>
<tr>
<td>Averted Gaze</td>
<td>0.9</td>
<td>26.8</td>
</tr>
<tr>
<td>Joint Attention</td>
<td>-25.6</td>
<td>29.1</td>
</tr>
</tbody>
</table>

With respect to imitation of Speech Rate, there was no main effect of experimental group (ASD, Control) on Speech Rate, $F(1,30) = 0.01$, $p = .945$ and no interaction effect between experimental group and Gaze Direction (Direct Gaze, Averted Gaze, Joint Attention), $F(2,60) = 0.05$, $p = .948$. There was a significant main effect of Gaze Direction, $F(2,60) = 3.23$, $p = .047$, partial eta squared $= .10$. Participants imitated the most in the condition with Direct Gaze and the least in the condition with Joint Attention. Pairwise comparisons with Bonferroni correction revealed a marginal difference between imitation in the condition with Direct Gaze and with Joint Attention, $t(37) = 1.97$, $p = .056$; there was no difference between Direct and Averted Gaze, $t(37) = 0.28$, $p = .783$, and between Averted Gaze and Joint Attention, $t(37) = 1.78$, $p = .083$. For descriptives, see Table 2.

2 In principle, other types of measures of vocal imitation are possible, such as mean pitch, (log) pitch range, or local minima and maxima. These values, however, are perceptually elusive and their role in vocal imitation is thus dubious; moreover, local minima and maxima are highly sensitive to the pitch tracker settings and thus not suitable for semi-automatic analysis.
Table 2: Descriptive values for Speech Rate Imitation (Hz) in the three experimental conditions (positive value indicates convergence to target’s values, negative value shows divergence).

<table>
<thead>
<tr>
<th>Group</th>
<th>EMM</th>
<th>S.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Direct Gaze</td>
<td>-1.03</td>
</tr>
<tr>
<td></td>
<td>Averted Gaze</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>Joint Attention</td>
<td>-2.37</td>
</tr>
<tr>
<td>Control</td>
<td>Direct Gaze</td>
<td>-0.98</td>
</tr>
<tr>
<td></td>
<td>Averted Gaze</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>Joint Attention</td>
<td>-2.07</td>
</tr>
</tbody>
</table>

Discussion and Conclusions

In this study, we explored vocal imitation of two prosodically independent features, pitch slope and speech rate, in relation to eye gaze direction of the model talker. Eye gaze direction and vocal mimicry are powerful nonverbal social signals (Postma-Nilsenová, Brunninkhuis, & Postma, 2013). Both pitch and speech rate provide important information regarding emotions, and attitudes of the speaker and their imitation convergence and facilitates feelings of empathy and affiliation. Results of several recent experimental studies indicated that imitation decreases in situations where gaze engagement is lacking (Wang, Newport, & Hamilton, 2011; Wang, Ramsey, & Hamilton, 2011; Postma-Nilsenová, Brunninkhuis, & Postma, 2013). This effect has, however, not been explored in contexts where the speaker initiates gaze following by first establishing eye contact and then directing her gaze towards a relevant object in the environment. Given that joint attention is an undeniable part of language learning and establishes the focus of gaze on imitated vocal patterns. The outcomes of our study, however, show that this is not the case and provide partial backing for the view that absence of direct gaze slows down and, possibly, disrupts the imitation process (Wang, Ramsey, & Hamilton, 2011). The process of establishing joint attention, as well as the capability to imitate nonverbal cues, are both considered to be problematic for individuals with Autism Spectrum Disorders (Vanvuchelen, Roeyers, & De Weerdt, 2011). In our experiment, we compared the performance of autistic speakers to a matched control group of typically functioning adults. We found no difference in the degree of vocal imitation between these groups, disregarding the gaze direction of the model talker. This finding is in accordance with other recent studies of imitation in adult autism (Slocombe et al., 2012) and can be interpreted in line with approaches that view imitation as an automatic process that is not necessarily absent in autism. According to these views, it is not imitation as such that lies at the core of empathy and perspective taking (Theory of Mind), but rather its inhibition (Santesteban et al., 2012). Future research should further explore the effect of gaze on imitation in contexts where gaze following is not initiated by the model talker but by the imitator who directs her attention. In particular, it is conceivable that imitation increases in contexts where the model talker “follows into” the focus of the speaker, in line with results previously reported for gestural and linguistic development. A setting in which the follow-into joint attention could be established would likely be more interactive; it remains to be seen if spontaneous imitation would result in the same patterns as those found in the current experiment. Another alternative experimental setup would be one where the object to which joint attention is directed is not additionally introduced into the visual scene (as in our experiment) but remains present in all conditions, thus not introducing a distraction that possibly disrupts the imitation process. Finally, it is possible that making use of emotionally charged stimuli (rather than emotionally neutral, as in our experiment) would increase the difference in prosodic imitation between ASD and typically functioning speakers.

Acknowledgements

The authors would like to thank the anonymous CogSci 2014 reviewers for their comments, to Susan Vermeer for her help with collecting experimental data, to Caroliene van Lier-Apps and Lilian van Ensel for facilitating the research, to Eric Postma and Paul Vogt for helpful suggestions, and to the participants in the experiment for their time.

References