Journal of Assistive Technologies

Recognizing voice: the child with autism spectrum disorder

Susan Jennifer Ni Chuileann Jean Quigley

Article information:

To cite this document:
Susan Jennifer Ni Chuileann Jean Quigley , (2016),"Recognizing voice: the child with autism spectrum disorder", Journal of Assistive Technologies, Vol. 10 Iss 3 pp. 140 - 152
Permanent link to this document:
http://dx.doi.org/10.1108/JAT-04-2015-0011

Peer-reviewed paper

Recognizing voice: the child with autism spectrum disorder

Susan Jennifer Ni Chuileann and Jean Quigley

Susan Jennifer Ni Chuileann and Jean Quigley are Lecturers, both at School of Psychology, Trinity College Dublin, Dublin, Ireland.

Abstract

Purpose – *This paper assesses the ability of the minimally verbal child with autism to recognise their own voice. The rationale for this study rests in recent advances in technology aimed at making the voice of speech generating devices (SGDs) sound more like the child using them (van Santen and Black, 2009). The purpose of this paper is to investigate the child's ability to actually recognise the sound of their own voice in a series of short experiments using computer-based methodology.*

Design/methodology/approach – *Using a voice-face matching computerised paradigm, the performance of 33 children with autism was compared to that of 27 children with developmental delay (DD), and 33 typically developing (TD) children. The children were matched for verbal and non-verbal ability and a training period was conducted prior to the main test to ensure children's understanding of what was expected of them.*

Findings – *The findings of this study suggest that the child with autism recognise the sound of their own voice at test, but with much greater difficulty than age-and-ability matched comparison groups. The implications of this finding are useful for researchers in the field of speech mimicry technology and manufacturers of SGD software packages. The paper also provides empirical insights about how the child with autism may process voice in their everyday social interactions.*

Research limitations/implications – *Some limitations to this study exist, for instance, there were only a small number of presentations involving self-voice in this task. This may have over simplified the process for the young TD children and the children with DD. Nevertheless, it is striking that despite being matched for non-verbal mental age, the children with autism performed significantly less well than either of the other two groups of children. However, future studies would benefit from adjusting the number of presentations of voice and face accordingly. It is also important to note that for some children with autism the simultaneous presentation of faces and voices may act more as an interference effect (Cook and Wilding, 1997; Joassin et al., 2004) than a facilitation effect (Molholm et al., 2002). Future studies may wish to test a subgroup on voice recognition without the aid of visual prompts.*

Practical implications – *The paper includes implications for the type of voice children with autism may prefer to use when communicating via a SGD. The authors suggest that if the child does not recognise or prefer the sound of their own natural voice on such devices, partial or complete abandonment of the SGD may occur.*

Originality/value – *This paper fulfils an identified need to research how children's abilities and preferences can be taken into account at the point of decision making for particular communication tools.*

Keywords *Voice, Speech generating devices, Typically developing child, Autism spectrum disorder, Communication tools, Developmental delay*

Paper type *Research paper*

Introduction

Received 10 April 2015
Revised 24 August 2015
23 October 2015
4 December 2015
Accepted 7 December 2015

In line with the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V; American Psychiatric Association, 2013), autism is characterised by difficulties with social interaction, behavioural flexibility and communication issues. The communication issues faced by verbal individuals with autism often are considered as rather unique to this disorder and include echolalia, prosodic oddities and rigid speech patterns (see Boesch *et al.*, 2013; Eigsti *et al.*, 2011; Waterhouse, 2013). A number of other communication issues associated with autism have been documented, such as late onset of speech (Eigsti *et al.*, 2011), a failure to respond satisfactorily to questions and comments (Capps *et al.*, 1998), pragmatic difficulties (Kelley *et al.*, 2006) and speech regression (Tager-Flusberg *et al.*, 2011). Moreover, there is evidence to suggest that up to 25 per cent of individuals with autism never gain any form of functional speech across the lifespan (Eigsti *et al.*, 2011; Tager-Flusberg *et al.*, 2011; Wing and Attwood, 1987). As such, the communication issues challenges faced by a number of individuals with autism suggest certain obstacles for full inclusion in their respective social worlds (Beukelman and Mirenda, 2012; Boesch *et al.*, 2013; Shane *et al.*, 2012).

Fortuitously, augmentative and alternative communication (AAC) strategies have been developed to help in dealing with these communication difficulties (Mirenda and Schuler, 1988; Shane *et al.*, 2012). Originally designed to replace (alternative) or enhance (augment) communication, AAC devices now range in technological sophistication from no-tech (i.e. sign language) to low-tech (i.e. picture exchange communication systems), to high-tech (i.e. speech generating devices (SGDs)) (Boesch *et al.*, 2013; Beukelman and Mirenda, 2012; Kee, 2012; Shane *et al.*, 2012).

Defined as "an electronic communication aide that produces digitised or synthesised speech upon activation by individuals" (Boesch *et al.*, 2013, p. 481), the SGD stands apart from other AAC as the speech output they produce allows the user to communicate verbally with anyone within earshot (see Boesch *et al.*, 2013; Sigafoos and Drasgow, 2001; van der Meer and Rispoli, 2010). Moreover, unlike other AAC, there is no expertise required on behalf of the listener to understand the message it produces (Boesch *et al.*, 2013). There are also suggestions that children with autism may adapt well to SGDs as these children are often attracted to technology and digital media (Dautenhahn, 1999; Goldsmith and LeBlanc, 2004; Kee, 2012).

One potential drawback to the SGD is that the voice on these devices rarely resembles the voice of the child using it (van Santen and Black, 2009). Significant strides are being made in the field of speech synthesis technology, however, with one particular system allowing SGD users with sufficient speaking ability to record speech for the purpose of developing their own synthetic voices as output (Bunnell *et al.*, 2004). For those SGD users who cannot speak, a Speaker Mimicry system has been created in which "only a small set of recordings is required from the SGD user (< 50 sentences) to adapt a high-quality synthetic voice based on a professional speaker to sound like their own" (Klabbers *et al.*, 2010, p. 2154). The underpinning rationale for these developments is that a device which sounds more like that of the child using it will psychologically reinforce powerful motivational factors and a sense of ownership for communication so that the frequency and richness of AAC use, and its acceptance by family members and friends, will be enhanced (van Santen and Black, 2009). This simple and yet stunningly creative innovation appears to make maximum sense given the role of voice as a marker of self and a signal for identification to others (see Belin *et al.*, 2004). But while those listening might better recognise this voice rich in accent, dialect, age, gender and ethnicity, it seems important to ascertain whether or not the child using the device can also.

Voice recognition is not trivial. Voice carries a wealth of information about the gender, age and emotional status of the speaker (Belin *et al.*, 2004). It is also well established that the ability to recognise individuals on the basis of their voice is developmentally sensitive and directly influenced by the stimulus set size, the duration of the voice sample and the listener's familiarity with the speaker's voice (Spence *et al.*, 2002). The typically developing (TD) child can recognise a number of people they are reasonably familiar with by the age of four, but in general they are much better at recognising faces than voices (Bartholomeus, 1973; Schuster, 1998). In relation to self-voice recognition, there is evidence to suggest that the typical four to eight year old children perform at near-adult level accuracy but with significant variation in their scores (Schuster, 1998; Strömbergsson, 2009). There is also some suggestion that girls recognise voices at a slightly more accurate rate than boys at test (Bartholomeus, 1973; Schuster, 1998; Strömbergsson, 2009).

Studies of voice recognition abilities by the individual with autism are less common than those conducted with the TD population. One recent study investigated the way adults with autism recognise and memorise vocal identity relative to adults with no diagnosis and found that while there was similar abilities to recognise voices, they did so in a way that was categorically different than the typical person (Lin *et al.*, 2015). Those with autism recognised voices on the basis of "exact acoustic features" rather than on the "acoustic patterns correlated to the speaker's physical and mental properties" (Lin *et al.*, 2015, p. 1).

Studies investigating voice-recognition in children with autism have focused on their lack of orientation to their own name (Osterling and Dawson, 1994) and their impaired performance on recognising emotionally expressive voices relative to other children with developmental delay (DD) (Hobson *et al.*, 1988). Two studies using voice-face matching paradigms with relatively able school-children with autism spectrum disorder (ASD) were conducted by Boucher *et al.* (1998, 2000). The first of these tasked the participant with matching the voice recordings of staff from the child's school to photographs of the same people at test (Boucher *et al.*, 1998). The findings of this study showed that the children with autism were impaired on this task relative to language-matched typically developing children. Significantly, the researchers noted that the children with autism had "rather more opportunities than controls for exposure to the adults whose voices and faces were used as stimuli given that they had been in school for longer than the controls" (Boucher *et al.*, 1998, p. 180). This led the researchers to conclude that seven to nine year old children with autism "either did not, or could not, utilise these opportunities to achieve normal familiarity" (Boucher *et al.*, 1998, p. 180). The second voice-face matching study was conducted with slightly older children with autism (Boucher *et al.*, 2000). This time no impairment was found in this group relative to language-matched children with specific-language impairment. As such, the findings from studies of familiar voice recognition in the child with autism are scarce, and provide mixed results.

Two main factors may contribute to this discrepancy in the studies conducted by Boucher *et al.* (1998, 2000). First, the faces and voices used as stimuli differed across the two studies with the voices and faces of adult staff used in the first study and a mixture of adult staff and pupils in the second study. Research suggests that recognition increases when the faces and voices of stimuli used as stimuli are similar in generational age to the participant (Rosa *et al.*, 2008). Second, the sample size used in the 1998 study was small, making generalisations of impairments noted contentious. There was also no typically developing (TD) group included in either the Boucher *et al.* (1998) or the Boucher *et al.* (2000) study, which potentially dilutes the possibility that the performance of the target group is autism specific or merely a function of general delayed development (Burack *et al.*, 2002). Studies of the ability of the child with autism to recognise their own voice appear to be non-existent to date and given the recent technological advances in SGDs for the child with autism, an investigation of children's recognition of and preferences for self-voice voice on SGDs seems timely and important.

One other potential factor implicated in the difficulties around voice perception and voice recognition is that of memory. A wealth of research shows that individuals with ASD experience some form of memory anomaly possibly from very early in life (see Boucher and Bowler, 2008, 2010). But the aspect of memory we are interested in is that of recognition memory. Yonelinas (1994) suggested that recognition could only really be understood as reflecting the contribution of two separable memory retrievable processes, namely, recollection and familiarity. Recollection can be considered as a kind of recall in which a recognised stimulus cues "recall of contextual information experienced within the episode in which the stimulus was encountered" (Bigham *et al.*, 2010, p. 879). Familiarity, on the other hand, refers to a conscious feeling that one has experienced a stimulus before without necessarily recalling any other information (Bigham *et al.*, 2010; Ni Chuileann and Quigley, 2012).

Accordingly, the first part of this study is a test of familiarity, based on the assumption that the child's recognition of the voices of already known people will be accomplished on foot of stimulus repetition, building on perceptual representations of people's faces and voices that may initially have been implicit (see Atkinson and Wescourt, 1975; Mandler, 1980; Boucher *et al.*, 2012).

Using photographs as cues, or prompts, the child's memory is tested in terms of the associations they have made in long-term memory for faces, names and voices (Mandler, 1980). The second part of this study arguably assesses children's recollection of personally relevant information from long-term memory. The child is tasked with recalling the recording of their voice from before and recognising it as their own.

Based on the literature, we anticipated that the child with autism would be impaired relative to age and ability matched groups of children with developmental delay (DD) and typically developing (TD) children in their ability to match the voices and faces of familiar people; children with autism would be impaired relative to age and ability matched groups of children with DD and TD children in their ability to match self-voice-to-self-face. We were interested as to whether non-verbal mental age (NVMA) would be associated with greater voice-face matching ability across participants. We were also interested to consider the potential influence of gender on participant's scores.

Method

Participants

Three groups took part in this study: a group of children ($n = 33$) with ASD with sufficient language to co-operate with formal testing (see Table I: all tables available at request via nichuils@tcd.ie); a group of TD children ($n = 33$) matched with the ASD group for verbal and nonverbal mental ages; and a group of children with DD without autism ($n = 27$) equated with the ASD group for chronological age (CA), verbal and nonverbal ability (see Table II). Baseline tests of verbal and nonverbal ability were the British Picture Vocabulary Scale (BPVS) (Dunn *et al.*, 1997).

All children in the ASD group were attending an autism-specific unit catering specifically for students with autism. Each class had a maximum of ten students. This unit was attached to a mainstream school. All had been diagnosed as autism by experienced psychiatrists and clinical psychologists using DSM-IV (APA, 1994) criteria, and all had scores of 30 or above on the Childhood Autism Rating Scale (Schopler *et al.*, 1988) as completed by independent educational psychologists. Recent assessments conducted by these psychologists showed that this group of children presented with an average IQ score of 50 as calculated via the Wechsler Abbreviated Intelligence Scale (WASI)(Wechsler, 1999). All had verbal quotients of less than 75, with verbal mental-age and NVMA equivalents of more than 5; 0 years on the BPVS.

Children in the TD group were attending the mainstream school attached to the autism-specific unit. These children had no record of autism and were described by their teachers as having no significant emotional, social or cognitive problems. This group of children were of average ability with IQ scores averaging at 97.8 on the WASI, and had verbal quotients on the BPVS of between 85 and 115.

The children in the DD group attended a special school catering specifically for students with additional needs. Previous assessments by educational psychologists noted that this group had no record of autism features of behaviour. All had verbal age equivalents of more than 50 years on the BPVS with an average IQ score of 54.9 as calculated via the WASI.

Informed consent to include participants in this research was obtained from parents or guardians via each child's school. Following baseline testing, each child was asked if they would like to continue working with the tester and they were clearly informed that they could refuse if they wished. In this way, assent was obtained from each child. No child declined to participate.

Design

This was a test designed primarily to test the ability of a child to match voice recordings of familiar people to their corresponding photographs (Study 1), andto match recordings of their own voice to their own photograph at test (Study 2). To do so, a forced-choice paradigm was used. This involved the design and presentation of a PowerPoint demonstration whereby each of the slides showed the photographs of two people the child knew from school and the voice of one of these individuals was played. The task for the child was to match the voice recording to the correct photograph. This PowerPoint also comprised photographs of the child themselves presented in fixed places throughout the presentation.

Materials and procedures for Studies 1 and 2

The children were tested on an individual basis in a familiar room in their school. For each session the tester and child sat beside each other at a table. Verbal input from the tester during testing was restricted to phrases such as "nearly finished" or "almost there" if any child's attention appeared to wander. A pre-prepared score sheet was always used.

For both parts of this study, a short training phase preceded the main test. The materials used in training were selected so as to pose no significant challenge to the child, rather the focus was on assessing the child's ability to understand the task demands *per se*, independent of their ability to succeed when presented with the actual main test material. Children were expected to reach specific criteria in order to be included at main test.

Training stimuli for Studies 1 and 2. Two sets of stimuli were prepared, one set of digital photographs and one set of audio recordings. For the digital stimuli, three standard colour photographs were taken of three pupils from the classrooms of the participants but not the participant themselves. The children were photographed against a neutral wall while looking into the camera. Their clothing was obscured by a white cape.

For the audio stimuli, three recordings were prepared. To this effect, each pupil whose face appeared in the training set was recorded via a hand-held digital recorder (www.phillips.com) saying "Hello! Can you point to my picture?" This particular speech sample was selected to maximise voice recognition cues by providing a sample longer than one second in duration (Bartholomeus, 1973; Bricker and Pruzansky, 1976) and to include both falling and rising intonation (Bartholomeus, 1973).

Using these visual and audio stimuli, a PowerPoint slideshow was prepared consisting of four slides in total. Each slide displayed two photographs presented side by side (e.g. A/B; A/C; B/C; B/A). The speech sample of just one of the two photographs shown was ever played on any presentation (e.g. the voice of A, C, B, and A).

Test stimuli for Studies 1 and 2. Two sets of test stimuli were prepared, one set of digital photographs and one set of audio recordings. For the digital stimuli, six standardised colour photographs, similar to the training photographs, but this time of five different pupils plus a photograph of the target participant. A recording of each child photographed (including that of the target child) saying "Hello! Can you point to my picture?" was also prepared.

Using these visual and audio stimuli, a PowerPoint slideshow was prepared, this time consisting of ten slides arranged in a predetermined format. The ten slides always contained the photographs of two children presented side-by-side as follows (1-2) (3-4), (5-6), (4-3), (2-4), (6-1), (3-1), (5-2), (4-6) and, (2-1).

The photograph of the target participant always corresponded to position 6. For each of the ten slides shown, the voice of just one person was played (e.g. 1, 4, 6, 3, 2, 1, 3, 5, 6 and 1). Directly after the ten slides were presented, they were presented again, with the voice of the other person played (e.g. 2, 3, 5, 4, 4, 6, 1, 2, 4 and 2). Therefore just as with every child whose face and voice was used at stimuli in this study, the photograph of the target child was shown on three occasions while the voice of the target child was played on two occasions. The rationale for this order of presentation and for having the target face act as a distractor on at least one occasion was to minimise the potential for false-positive identification if the target child's photo was always accompanied by their voice.

Procedure for Studies 1 and 2. The training and testing were completed in a single session. The aim of the practice phase was to ensure children's understanding of the task of matching a recording of voice to a photograph via a PowerPoint demonstration. The training procedure was as follows.

The participants were invited to sit at a desk to the right of the tester. In the centre of the table, facing the participant and the tester, a laptop with the training and main test PowerPoint was located. The tester put up the first slide showing (A-B) side-by side and pointing at photograph "A" named the person and said, "You know 'A'. 'A' is in your class". After a pause of approximately three seconds, the tester pointed to photograph "B" named that person, and said "You know 'B'. 'B' is in your class". This procedure was repeated for the remaining slides.

The tester then drew the child's attention back to the screen and returned to the start of the practice phase PowerPoint and said "I'm going to play a recording of a voice next. It will be the voice of one of these people you just saw. I want you to point to the person you think the voice belongs to. Will you do that for me?" The tester then activated the voice of A. If the child pointed to A the next slide showing C and D side-by-side was presented and the next recording (i.e. voice C) was played. If the child chose the incorrect photograph the tester said "I think it must be this person" and moved onto the next slide. This procedure was repeated using a different order of voices until each of the four slides (e.g. A-B, C-D, A-D, and B-C) had been played twice and participants had made four consecutive correct responses within the eight presentations. On average, this session lasted approximately seven minutes per child. All children met this criterion and progressed to the main test.

For the main test the child was told there were lots more photographs to see and were asked if they wished to continue. All children were content to progress to the main test. The procedure was the same as used at practice whereby the child was shown the ten slides and the child was helped to name the people on the screen to ensure that the children recognised the pupils and the photo of themselves.

The PowerPoint slideshow was then played again but this time the tester said "I'm going to play a recording of a voice next. It will be the voice of one of these people you just saw. I want you to point to the person you think the voice belongs to. Will you do that for me?" The first slide was shown with the photographs of two individuals presented in a side-by-side format and the speech sample of one of the individuals was played. Each of the ten slides was shown for a total of five seconds as this time corresponded with the length of time it had taken to name each photographed person previously. This procedure was repeated until all ten slides were shown.

A score of 1 was awarded for all correct matches and the next slide was presented. If the child failed to respond (more than ten seconds), or made an incorrect response, a score of 0 was awarded. Scores thus ranged from 0 to 20.

Results of Study 1

In the first instance, we anticipated that children with autism would show impairments in their ability to match familiar voices to familiar faces compared to comparison groups. We further expected to find that higher CA, IQ, VMA, (and possibly gender), would correlate with greater voice-face matching and self-voice recognition in all three groups of children.

It was clear that some participants ($n=8$; ASD = 6; DD = 2) scored at chance (scores of 10 or below) when matching familiar-voices to familiar faces during this study (see Table III).

A series of independent samples t-tests were conducted to compare the CA, NVMA, VMA and IQ of the six children who scored below chance with the 27 who achieved higher scores in the ASD group. There was no significant difference in CA scores between the six children who scored below chance ($M=9.3$, SD = 2.4) and the remaining 27 children, $M=11.0$, SD = 3.1, t (31) = 1.21, $p=0.23$ (two-tailed). Nor was there any significant difference in terms of NVMA between the six children ($M=4.9$, SD = 1.5) and the 27 children, $M=5.3$, SD = 1.7, t (30) = 0.53, $p=0.59$ (two-tailed). There was no significant difference in relation to the VMA of the six children ($M=3.5$, SD = 2.0) and the remaining 27, $M=4.0$, SD = 1.8, t (31) = –0.49, $p=0.62$ (two-tailed). Finally, there was no significant difference in IQ between the six children ($M=52.8$, SD = 9.2), and those with ASD who scored above chance, $M=48.35$, SD = 14.95, t (31) = 0.55,

$p = 0.49$ (two-tailed). Therefore, to ensure the validity of the overall findings, it was decided to remove these eight children from the data set prior to conducting any further analysis leaving a remainder of 27 participants with autism, 25 participants with DD and the full set of 33 TD participants (see Table IV).

Although apparent that some children with autism in this study ($n = 4$) achieved maximum scores in this particular task, in the main these children successfully matched familiar voices to familiar faces less accurately than did their ability-matched TD counterparts ($n = 21$). Numerically, they also performed less well than age-and-ability matched children with DD ($n = 10$) but this difference did not reach statistical significance.

Diagnosis and matching ability

A one-way between groups analysis of variance (ANOVA) was conducted to explore the effect of diagnosis (ASD/DD/TD) on participants' mean scores of voice-face matching and this showed a statistically significant difference in voice-face matching scores for the three groups: $F\ (2, 84) = 11.72$, $p < 0.0005$. *Post hoc* comparisons using the Tukey HSD test indicated that the mean score of the participants with autism for matching voices to faces ($M = 15.85$, $SD = 2.97$) was significantly lower than that of the ability matched younger TD participants on this task ($M = 18.93$, $SD = 1.57$). This difference in group means was moderate, with the effect size calculated using η^2 as 0.22 (Cohen, 1992, p. 157). As noted previously, the participants with DD ($M = 17.37$, $SD = 2.81$) did not differ significantly from either the ASD or the TD groups on this task. Consequently the results of this ANOVA support our hypothesis that the ability to match the voices of highly recognisable people to photographs of their faces was most difficult for the children with ASD.

NVMA and matching ability

While the three groups of children were matched on mean average NVMA, the range of NVMA scores across participants was from 38 to 103 months. To consider the potential impact of NVMA on overall scores, therefore, we conducted a two-way between-groups ANOVA. The children were divided into three groups on the basis of their NVMA (Group 1: $n = 28$, 62 months or less; Group 2: $n = 29$, 63-72 months; Group 3: $n = 27$, 73 months plus). The main effect for NVMA, $F\ (2, 75) = 0.74$, $p = 0.48$, did not reach statistical significance and the interaction effect between diagnosis (ASD, DD, TD) and NVMA was not statistically significant, $F\ (4, 75) = 0.75$, $p = 0.55$.

Gender and matching ability

A second two-way ANOVA was conducted to explore the potential influence of gender and diagnosis on face-voice matching scores. The interaction effect between gender and diagnosis was not statistically significant $F\ (2, 79) = 0.74$, $p = 0.47$. There was however a statistically significant main effect for gender, $F\ (1, 79) = 7.95$, $p = 0.006$, and the effect size was moderate (partial $\eta^2 = 0.006$). As such, males with ASD matched voices to faces ($M = 16.23$, $SD = 2.5$) with greater accuracy than females with ASD ($M = 15.1$, $SD = 3.7$). This gender divide was similar for males with DD ($M = 18.1$, $SD = 2.5$) compared to females with DD ($M = 15.6$, $SD = 3.2$), and for TD males ($M = 19.\ 3$, $SD = 1.1$) as opposed to TD females ($M = 18.35$, $SD = 1.9$).

Results of Study 2

In this part of the study we used the same two-choice forced recognition paradigm utilised in our face-voice matching study to assess the ability of the children with ASD to match a recording of their own voice to that of their own photograph at test. We compared the performance of the 33 children with ASD with performance in a group of young TD children matched with the ASD group for NVMA and an age matched group of children with DD also matched with the ASD group on NVMA.

Self-voice-self-face matching in ASD has not previously been previously tested but on the basis of findings from voice-face matching it was predicted that: children with ASD would be impaired relative to age and ability matched groups of children with DD and TD children; that girls would match self-voice to self-faces more accurately than boys; and given the range of NVMA scores in the two clinical groups, higher NVMA scores would be associated with greater self-voice-self-face matching ability.

Diagnosis and matching ability

A one-way between-groups ANOVA was conducted to explore the impact of diagnosis (ASD, DD, TD) on participant ability to match recordings of self-voice to self-image at test. There was a statistically significant difference at the $p < 0.05$ level in scores for the three groups: $F\ (2, 88) = 16.6$, $p = 0.0005$, with the difference in mean scores between the groups calculated at 0.2 using η^2. *Post hoc* comparisons using the Tukey test indicated that the mean score for the children with ASD ($M = 0.878$, SD = 0.92) was significantly different from the children with DD ($M = 1.60$, SD = 0.70) and the TD children ($M = 1.84$, SD = 0.36). While reaching numerical difference, there was no statistically significant difference between the children with DD and the TD group (see Table V).

NVMA and matching ability

Despite the three groups of children being matched on mean average NVMA, the NVMA scores across participants ranged from 38 to 103 months. A one-way ANOVA was then conducted to consider the effect of NVMA on participant ability to match self-voice recordings to self-image at test. The children were divided into three equal-sized groups according to their NVMA (Group 1: 62 months or less; Group 2, 63-72 months; Group 3, 73+ months). There was no statistically significant difference at the $p = 0.05$ level on this test for the three age groups: $F\ (2, 88) = 2.3$, $p = 0.10$.

Gender and matching ability

An independent-samples *t*-test was conducted to compare participant ability to match recordings of self-voice to self-image at test. There was no significant difference in the scores for boys ($M = 1.48$, SD = 0.80) and girls, $M = 1.34$, SD = 0.83; $t\ (89) = 0.788$, $p = 0.43$ (two-tailed). The magnitude of difference between the means (mean difference = 0.13, 95 per cent CI: −0.21-0.49) was very small ($\eta^2 = 0.006$).

Discussion

A number of individuals with ASD cannot communicate functionally by voice alone (see Eigsti *et al.*, 2011; Waterhouse, 2013). Developments in the field of assistive technology mean that a number of voice enabled AAC devices, also called SGDs, have become popular and valuable communication tools for this group. Recently, speaker synthesis technology has emerged to make the voice on the regular SGD sound more like the child using it. This innovative development heralds the possibility of the child with autism identifying more with the SGD, which may enhance the frequency and richness of their AAC use.

Research into voice perception in autism suggest that they may take longer to orient to vocal stimuli than TD children (Osterling and Dawson, 1994), and they may be less able than TD individuals to recognise the emotion expressed in voice (Hobson *et al.*, 1988). There is also some evidence that even when photographs are used as prompts to cue memory for the people whose voices are being used as test stimuli, that the child with autism shows less accuracy than either children with learning difficulties or TD children (Boucher *et al.*, 1998, 2000). If the verbally challenged child with ASD demonstrates impoverished ability to recognise the voice of people familiar to them, will they recognise and prefer the sound of their own recorded voice on SGDs? With this aim, this study tested familiar and self-voice recognition in a group of children and adolescents with ASD, and ability-matched group of young TD children, and an age and ability-matched group with learning difficulties.

Our overall finding is that voice recognition in general is most difficult for the children and adolescents with ASD. Neither the gender nor the NVMA of the child made any significant difference to their overall accuracy at matching vocal stimuli to visual stimuli at test. This finding has both theoretical and practical implications for the field of research dedicated to assistive technologies.

From a theoretical standpoint, the findings provide further evidence of impoverished recognition memory in children and adolescents with ASD (see Bigham *et al.*, 2010; Boucher and Bowler, 2010; Ni Chuileann and Quigley, 2012). For instance, Study 1 was a test of familiarity, with the photographs serving as visual prompts of the speaker's identity in the child's long-term memory. Previous research suggests this ability is at adult-level performance by the age of four in TD children (Bartholomeus, 1973). In this instance however, most children with ASD were significantly less able to match voices to faces than either children with DD or young TD children.

This was an important finding given that the stimuli used at test comprised the faces and voices of the teachers and children from the children's respective autism-specific schools which should therefore, have been highly familiar. Implicit in this finding is a suggestion that relational encoding may be impaired in the child with ASD (Bowler, 2011). In other words, children with autism may not be creating implicit associations between the faces, voices, names and other social information linking person-specific information together in memory. If correct, this finding is consistent with previous suggestions implicating a combination of impaired processing of complex information with intact processing of simple information (Boucher *et al.*, 2012; Minshew *et al.*, 2002). Indeed, just four children with ASD achieved full scores on this test of familiar voice-face matching.

Recognition memory is also underpinned by recollection (Yonelinas *et al.*, 2010). The sight and sound of the child's own voice and face at test have served as recognised stimulus stored in long-term memory. For the TD children, this did occur, with almost ninety per cent of this group recognising their own voice. Additionally, as self-recognition is often compromised in atypical populations (Christiana *et al.*, in press; Rosa *et al.*, 2008) it was unsurprising to note that one third of the children with DD did not succeed on this task. Significantly, however, just 12 children with ASD matched their own voice to their own photograph at test.

From a practical standpoint, AAC manufacturers and those involved in speech synthesis technology development may need to consider the child's abilities and preferences for personalised voices on SGDs for individuals with autism. Personalised SGDs may indeed enhance the communications of some children with autism using them (van Santen and Black, 2009), but for others, the sound of their own voice may not be recognised or preferred. While speech synthesis technology is innovative and impressive in the extreme, it may be wise not to be too "dazzled 'by what is now available, and to instead ensure methodological, clinical processes are conducted so that each child is matched with the most optimal communication technology that suits them (Shane *et al.*, 2012, p. 1229; see also Blischak and Ho, 2000; Shane and Bashir, 1980).

Some limitations to this study exist, for instance, there was a very small number of presentations involving self-voice in the overall task. This may have over simplified the process for the young TD children and the children with DD. Moreover, because ASD is characterised by severe learning and language difficulties, these children may not be used to speaking often or for long periods. Thus, their familiarity with a recording of their self-voice may have been impoverished (see Spence *et al.*, 2002). Second, and in a related vein, due to their low language production, it is necessary to ask them to repeat the words for subsequent recognition after the researcher and to later edit out the researcher's voice. Editing children's recordings has meant that phonologically challenged children could not recognise their own recorded voice in a previous study (Schuster, 1998). In the event, children with ASD were significantly impaired relative to controls, but 12 children with autism did succeed on this task. It is also the case that for some children, the simultaneous presentation of faces and voices may have acted as an interference effect (Cook and Wilding, 1997; Joassin *et al.*, 2004) than a facilitation effect (Molholm *et al.*, 2002). Future studies may wish to test a subgroup on voice recognition tasks minus the aid of visual stimuli.

Overall, this study is useful as it suggests that for a number of children with autism, a personalised voice on a SGD may enhance the frequency and quality of their use of the device to communicate, as they may well identify with that voice and experience more of a "sense of owness" for their communications (van Santen and Black, 2009). These are exciting times for the child with autism living in the twenty-first century. After all, like all children, the child with autism is a digital native, born into a world of technology and communicative platforms (Kee, 2012). It remains to be seen how they will comprehend and traverse their social worlds with the assistance of personalised SGDs, but the findings of this study suggest that for some, it may well be very successful.

Future researchers in the field of assistive technologies may choose to investigate the advantages of self-voice on SGDs for children with autism with larger sample sizes. It would also be of interest to consider the reactions of children with autism of more varied ages, cultures and abilities. Children's preferences for speech output on AAC is after-all, one of the most under researched areas in the field of assistive technologies (van der Meer and Rispoli, 2010). It would also be of considerable interest to test children's memory for self-related material such as their voice, face and actions as important variables to be factored into ongoing analysis of voice-recognition on SGDs in this group of children (see Boucher and Bowler, 2010).

Glossary

BPVS	British Picture Vocabulary Scale
CA	chronological age
DD	developmental delay
ASD	autism spectrum disorder
NVMA	non-verbal mental age
TD	typically developing
VMA	verbal mental age
WAIS	Wechsler Abbreviated Intelligence Scale

References

American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V), 5th ed., APA, Washington, DC.

Atkinson, R.C. and Wescourt, K.T. (1975), "Some remarks on a theory of memory", in Rabbitt, P.M.A. and Dornic, S. (Eds), *Attention and Performance V*, Academic Press, London, pp. 101-46.

Bartholomeus, B. (1973), "Voice identification by nursery school children", *Canadian Journal of Psychology*, Vol. 27, pp. 464-72.

Belin, P., Fecteau, S. and Bédard, C. (2004), "Thinking the voice: neural correlates of voice perception", *Trends in Cognitive Science*, Vol. 3, pp. 129-35.

Beukelman, D.R. and Mirenda, P. (2012), *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs*, Paul H. Brookes Publishing Co, Baltimore, MD.

Bigham, S., Boucher, J., Mayes, A. and Anns, S. (2010), "Assessing recollection and familiarity in autistic spectrum disorders: methods and findings", *Journal of Autism and Developmental Disorders*, Vol. 40, pp. 878-89.

Blischak, D.M. and Ho, K.M. (2000), "School-based augmentative and alternative communication evaluation reports", *Contemporary Issues in Communication Science and Disorders*, Vol. 27, pp. 70-81.

Boesch, M., Wendt, O., Subramanian, A. and Hsu, N. (2013), "Comparative efficacy of the picture exchange communication system (PECS) versus a speech-generating device: effects on requesting skills", *Research in Autism Spectrum Disorders*, Vol. 7, pp. 480-93.

Boucher, J. and Bowler, D. (2010), *Memory in Autism: Theory and Evidence*, Cambridge University Press, Cambridge.

Boucher, J., Lewis, V. and Collis, G.M. (1998), "Familiar face and voice matching and recognition in children with autism", *Journal of Child Psychology and Psychiatry*, Vol. 39, pp. 171-81.

Boucher, J., Lewis, V. and Collis, G.M. (2000), "Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children", *Journal of Child Psychology and Psychiatry*, Vol. 41, pp. 847-57.

Boucher, J., Mayes, A. and Bigham, S. (2012), "Memory in autistic spectrum disorder", *Psychological Bulletin*, Vol. 138 No. 3, pp. 458-96.

Bowler, D.M. (2011), *Memory in Autism: Interview with Dermot Bowler*, James Brooks, available at: http://jimbrux.wordpress.com

Bricker, P.D. and Pruzansky, S. (1976), "Speaker recognition", in Lass, N.J. (Ed.), *Contemporary Issues in Experimental Phonetics*, Academic, New York, NY, pp. 295-326.

Bunnell, H., Gray, J., Pennington, C. and Yarrington, D. (2004), "Automatic construction of concatenative speech synthesis databases for AAC", *American Speech Language and Hearing Association Conference, Philadelphia, PA*.

Burack, J.A., Iarocci, G., Bowler, D. and Mottron, L. (2002), "Benefits and pitfalls in merging disciplines", *Development and Psychopathology*, Vol. 14, pp. 225-37.

Capps, L., Kehres, J. and Sigman, M. (1998), "Conversational abilities among children with autism and children with developmental delays", *Autism*, Vol. 2, pp. 325-44.

Christiana, W., Malcolm, S., Johnson, A. and Keenan, J.P. (in press), "Mirror self-recognition in pervasive developmental disorders", *Social Neuroscience*.

Cohen, J. (1992), "A power primer", *Psychological Bulletin*, Vol. 112 No. 1, pp. 155-9.

Cook, S.A. and Wilding, J.M. (1997), "Earwitness testimony 2: voices, faces and context", *Applied Cognitive Psychology*, Vol. 11, pp. 527-41.

Dautenhahn, K. (1999), "Robots as social actors: aurora and the case of autism", *Proceedings CT99, The Third International Cognitive Technology Conference, San Francisco, CA*, *August*, pp. 359-74.

Dunn, L.M., Dunn, L.M., Whetton, C. and Burley, J. (1997), *The British Picture Vocabulary Scale (BPVS)*, 2nd ed., National Foundation for Educational Research, Windsor.

Eigsti, I.M., de Marchena, A.B., Schuh, J.M. and Kelley, E. (2011), "Language acquisition in autism spectrum disorders: a developmental review", *Research in Autism Spectrum Disorders*, Vol. 5 No. 2, pp. 681-91.

Goldsmith, T.R. and LeBlanc, L.A. (2004), "Use of technology for children with autism", *Journal of Early and Intensive Behavioural Intervention*, Vol. 1, pp. 166-78.

Hobson, R.P., Ouston, J. and Lee, T. (1988), "Emotion recognition in autism: coordinating faces and voices", *Psychological Medicine*, Vol. 18, pp. 911-23.

Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M. and Campanella, S. (2004), "When audition alters vision: an event related potential; study of the cross-modal interactions between faces and voices", *Neuroscience Letters*, Vol. 369 No. 2, pp. 132-7.

Kee, N.K.N. (2012), "Notes on digital wisdom in the 21st century", available at: www.keefamily.org/announcement/Norman_Submission.pdf (accessed 11 September 2013).

Kelley, E., Paul, J.J., Fein, D. and Naigles, L.R. (2006), "Residual language deficits in optimal outcome children with a history of autism", *Journal of Autism and Developmental Disorders*, Vol. 36, pp. 807-28.

Klabbers, E., Kain, A. and van Santen, J. (2010), "Evaluation of speaker mimic technology for personalizing SGD voices", *Proceedings Interspeech 2010, Makuhari*, pp. 2154-7.

Lin, I.-F., Yamada, T., Komine, Y., Kato, N., Kato, M. and Kashino, M. (2015), "Vocal identity recognition in autism spectrum disorder", *PLoS ONE*, Vol. 10 No. 6, p. e0129451, doi: 10.1371/journal.pone.0129451.

Mandler, G. (1980), "Recognizing: the judgment of previous occurrence", *Psychological Review*, Vol. 87 No. 3, pp. 252-71.

Minshew, N.J., Meyer, J. and Goldstein, G. (2002), "Abstract reasoning in autism: a dissociation between concept formation and concept identification", *Neuropsychology*, Vol. 16, pp. 327-34.

Mirenda, P. and Schuler, A. (1988), “Augmenting communication for persons with autism: issues and strategies”, *Topics in Language Disorders*, Vol. 9, pp. 24-43.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E. and Foxe, J.J. (2002), “Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study”, *Brain Research. Cognitive Brain Research*, Vol. 14, pp. 115-28.

Ni Chuileann, S. and Quigley, J. (2012), “Assessing recollection and familiarity in low-functioning autism”, *Journal of Autism and Developmental Disorders*, Vol. 43 No. 6, pp. 1406-22.

Osterling, J. and Dawson, G. (1994), “Early recognition of children with autism: a study of 1st birthday home videotapes”, *Journal of Autism and Developmental Disorders*, Vol. 24, pp. 247-57.

Rosa, C., Lassonde, M., Pinard, C., Keenan, J.P. and Belin, P. (2008), “Investigations of hemispheric specialization of self-voice recognition”, *Brain & Cognition*, Vol. 68, pp. 204-14.

Schopler, E., Reichler, R. and Rochen Renner, B. (1988), *The Childhood Autism Rating Scale*, Western Psychological Services.

Schuster, L.I. (1998), “The perception of correctly and incorrectly produced /r/”, *Journal of Speech, Language, and Hearing Research*, Vol. 41, pp. 941-50.

Shane, H. and Bashir, A.S. (1980), “Election criteria for the adoption of an augmentative communication system: preliminary considerations”, *Journal of Speech and Hearing Disorders*, Vol. 45, pp. 408-14.

Shane, H.C., Laubscher, E.H., Schlosser, R.W., Flynn, S., Sorce, J.F. and Abramson, J. (2012), “Applying technology to visually support language and communication in individuals with autism spectrum disorders”, *Journal of Autism and Developmental Disorders*, Vol. 42 No. 6, pp. 1228-35.

Sigafoos, J. and Drasgow, E. (2001), “Conditional use of aided and unaided AAC: a review and clinical case demonstration”, *Focus on Autism and Other Developmental Disabilities*, Vol. 16, pp. 152-61.

Spence, M.J., Rollins, P.R. and Jerger, S. (2002), “Children’s recognition of cartoon voices”, *Journal of Speech, Language, and Hearing Research*, Vol. 45, pp. 214-22.

Strömbergsson, S. (2009), “Self-voice recognition in 4 to 5 year-old children”, *Proceedings, FONETIK, Department of Speech, Music and Hearing*, available at: www.speech.kth.se/prod/publications/files/3385.pdf (accessed 4 September 2013).

Tager-Flusberg, H., Edelson, L. and Luyster, R. (2011), “Language and communication in autism spectrum disorders”, in Amaral, D.G., Dawson, G. and Geschwind, D.H. (Eds), *Autism Spectrum Disorders*, Oxford University Press, New York, NY, pp. 1-21.

van der Meer, L.A. and Rispoli, M. (2010), “Communication interventions involving speech-generating devices for children with autism: a review of the literature”, *Developmental Neurorehabilitation*, Vol. 13 No. 4, pp. 294-306.

van Santen, J. and Black, L. (2009), “In your own voice: personal augmentative and alternative communication voices for minimally verbal children with autism spectrum disorders”, *Proceedings Interspeech 2010, Makuhari*.

Waterhouse, L. (2013), *Rethinking Autism: Variation and Complexity*, 1st ed., Elsevier and Academic Press, San Diego, CA.

Wechsler, D. (1999), *Wechsler Abbreviated Scales of Intelligence*, Psychological Corporation, San Antonio, TX.

Wing, L. and Attwood, A.J. (1987), “Syndromes of autism and atypical development”, in Cohen, D.J., Donellan, A. M. and Paul, R. (Eds), *Handbook of Autism and Pervasive Development Disorders*, Wiley, New York, NY, pp. 3-19.

Yonelinas, A.P. (1994), “Receiver-operating characteristics in recognition memory: evidence for a dual-process model”, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, Vol. 20, pp. 1341-54.

Yonelinas, A.P., Aly, M., Wang, W.C. and Koen, J.D. (2010), “Recollection and familiarity: examining controversial assumptions and new directions”, *Hippocampus*, Vol. 20 No. 11, pp. 1178-94.

Further reading

Ellis, K. (1990), *Autism: Professional Perspectives and Practice*, Chapmen and Hall, London.

Jacoby, L.L. (1991), “A process dissociation framework: separating automatic from intentional uses of memory”, *Journal of Memory and Language*, Vol. 30 No. 5, pp. 513-41.

Ozonoff, S., Pennington, B.F. and Rogers, S.J. (1990), "Are there emotion perception deficits in young autistic children?", *Journal of Child Psychology and Psychiatry*, Vol. 31, pp. 343-61.

Pollack, I., Pickett, J.M. and Sumby, W.H. (1954), "On the identification of speakers by voice", *Journal of the Acoustic Society of America*, Vol. 26, pp. 403-6.

Wing, L. (2001), *The Autistic Spectrum: A Parents' Guide to Understanding and Helping Your Child*, Ulysses Press, Berkeley, CA.

Corresponding author

Susan Jennifer Ni Chuileann can be contacted at: nichuils@tcd.ie