Cognition 153 (2016) 1–5

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Similar impressions of humanness for human and artificial singing voices in autism spectrum disorders

Shinji Kuriki [a,1], Yuri Tamura [a,1], Miki Igarashi [b], Nobumasa Kato [b], Tamami Nakano [a,c,*]

[a] Department of Brain Physiology, Osaka University School of Medicine, Osaka, Japan
[b] Department of Psychiatry, Showa University School of Medicine, Tokyo, Japan
[c] Graduate School of Frontiers Biosciences, Osaka University, Osaka, Japan

ARTICLE INFO

Article history:
Received 16 November 2015
Revised 30 March 2016
Accepted 5 April 2016

Keywords:
Autism
Artificial voice
Humanness
Music

ABSTRACT

People with autism spectrum disorder (ASD) exhibit impairments in the perception of and orientation to social information related to humans, and some people with ASD show higher preference toward human-like robots than other humans. We speculated that this behavioural bias in people with ASD is caused by a weakness in their perception of humanness. To address this issue, we investigated whether people with ASD detect a subtle difference between the same song sung by human and artificial voices even when the lyrics, melody and rhythm are identical. People without ASD answered that the songs sung by a human voice evoked more impressions of humanness (human-likeness, animateness, naturalness, emotion) and more positive feelings (warmth, familiarity, comfort) than those sung by an artificial voice. In contrast, people with ASD had similar impressions of humanness and positive feelings for the songs sung by the human and artificial voices. The evaluations of musical characteristics (complexity, regularity, brightness) did not differ between people with and without ASD. These results suggest that people with ASD are weak in their ability to perceive psychological attributes of humanness.

1. Introduction

Autism spectrum disorder (ASD) is a developmental disorder characterized by difficulties in social cognition and communication. People with ASD exhibit impairments in orientating to social stimuli, such as human faces (Jemel, Mottron, & Dawson, 2006; Nakano et al., 2010) and speech (Alcantara, Weisblatt, Moore, & Bolton, 2004; Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998). Instead, some of them show a strong affinity to robots and computers (Diehl, Schmitt, Villano, & Crowell, 2012). A previous study also reported that children with ASD automatically imitated the behaviour of a robot much better than they did that of a human (Pierno, Mari, Lusher, & Castiello, 2008). With regards to this preference for artificial objects rather than humans in autism, Baron-Cohen suggested that individuals with autism are attracted to systems of low variance (such as machines) and less sensitive to systems where there is maximal variance as is the case with human behaviour (Baron-Cohen, 2006).

In general, a human feels a stronger affinity toward another human than they do toward artificial objects. In particular, the subtle differences between a real human and a human-like object evoke an impression of a substantial qualitative difference between them that is accompanied by a negative feeling toward the human-like object. This is a well-known phenomenon known as the "uncanny valley phenomenon" (Mori, 1970). Research verifying the uncanny valley phenomenon clearly demonstrates that people perceive not only a physical difference but also a qualitative difference between humans and artificial objects. Haslam proposed that people have a unique sense of humanness, a special sense of "human nature," that involves emotion, warmth, and cognitive flexibility and is opposed to "mechanistic" dehumanization (Haslam, 2006). Thus, the dissociation between physical and psychological attributes of humanness evokes a feeling of repulsion as manifested by the "uncanny valley phenomenon". In contrast, the individuals with ASD do not show a preference bias toward human (Nakano et al., 2010) nor a repulsion toward the artificial objects rather than real humans (Diehl et al., 2012). We speculated that individuals with ASD face difficulties in perceiving the psychological aspects of humanness from human-related information, and thus they were unable to detect subtle differences between human and human-like objects. Therefore, they did not

* Corresponding author at: Graduate School of Frontiers Biosciences, Osaka University, 1-3, Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: tamami_nakano@fbs.osaka-u.ac.jp (T. Nakano).
[1] These authors equally contributed to this study.

http://dx.doi.org/10.1016/j.cognition.2016.04.004
0010-0277/

demonstrate any repulsion toward the artificial objects regardless of its degree of human resemblance. To test this hypothesis, the present study examined the impressions of humanness and feeling experienced by people with ASD upon hearing the same song sung by a real human voice and an artificial voice. Although many previous studies have explored the perception of humanness using humanoid robots or computer-generated 3D animation (such as Broadbent et al., 2013; Gray & Wegner, 2012; Thompson, Trafton, & McKnight, 2011), it is technically difficult to create artificial objects with the appearance and movements resembling those of an actual human. To address this problem, we focused on creating an artificial singing voice because the technology for synthesizing an artificial voice has advanced to a point that it can be used in daily life. Our previous study used the same auditory stimuli and demonstrated that a human singing voice evoked a sense of humanness and positive emotion much more strongly than an artificial singing voice did in typically developing adults, even though the lyrics, melody, and rhythm were identical (Tamura, Kuriki, & Nakano, 2015). We expected that if people with ASD had a weakness in their sense of humanness, they would not detect a subtle difference between the stimuli and would have similar impressions of humanness and positive feelings for the human and artificial singing voices.

2. Method

2.1. Participants

Fourteen adults with ASD (10 male, 4 female; mean age, 27.6 years; range, 20–39 years) and fourteen adults without ASD (10 male, 4 female; mean age, 28.9 years; range, 21–37 years) participated in this study. The two groups were matched for sex and approximate age. ASD diagnoses were based on the clinical judgment of medical specialists according to the DSM-V criteria. All ASD participants were high functioning, with their full-scale IQ (Wechsler Adult Intelligence Scale, 3rd edition, WAIS-III) exceeding 89 (mean: 108.9, range: 89–120). The mean verbal IQ was 113.2 (range: 82–136), and the mean performance IQ was 101.4 (range: 82–114). All participants completed the Japanese version of the autism spectrum quotient (AQ) test (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The mean AQ score was 37.0 (range: 24–46) in the ASD group and 15.3 (range: 10–23) in the control group. The participants were chosen regardless of their musical experience. Before the experiment, the participants were asked about their musical preferences to ensure that they were not familiar with artificial voices in music. The study was approved by the review board of Osaka University, and all participants gave written informed consent before participation.

2.2. Auditory stimuli

We used the same 12 songs that were used in our previous study (Tamura et al., 2015). These songs were sung by an artificial voice created using the Vocaloid software (Yamaha Corporation, Japan) and were also covered by human singers. Vocaloid is a software package with a library that contains actual human voice samples of all phonemes in a given language (Japanese or English, depending on the library type). All songs were in a female voice with Japanese lyrics. We selected a 15-s segment that contained lyrics from each song and created a total of 24 WAVE files (sampling frequency, 44,100 Hz), each of which lasted 15 s, with a tapering effect on the first and last second. The average (root mean square) power was adjusted such that there was no difference in power between a human and an artificial voice singing the same song. The comparison of acoustic features between a human and artificial singing voices was described in our previous study (Tamura et al., 2015).

2.3. Procedure

Participants sat in front of a pair of speakers (Computer MusicMonitor, BOSE), through which the auditory stimuli were presented. Each of the 24 song segments was presented once and in random order, although a song sung by the human voice did not subsequently follow the same song in the artificial voice, and vice versa. The participants were not informed in advance that artificial singing voices were included in the stimuli.

After each song was presented, the participant completed a questionnaire in which he or she rated the song on a scale of 1–5 for ten items from three categories. The first category is the "human nature" dimension of humanness, which contrasts humans with machines and automata as proposed by Haslam (2006). The participants evaluated humanness (1 = mechanistic, 5 = human-like), emotion (1 = lack of emotion, 5 = rich in emotion), animateness (1 = inanimate, 5 = animate), and naturalness (1 = unnatural, 5 = natural). The second category is the internal positive feelings elicited by the singing voice. As demonstrated by the uncanny valley phenomenon, people generally experience positive feelings toward actual humans but not robots. Thus, we expected that the evaluations of positive feelings would correlate with the evaluations of humanness. The participants were asked to rate familiarity (1 = unfamiliar, 5 = familiar), warmth (1 = cold, 5 = warm), and comfort (1 = irritating, 5 = comfortable). The third category is related to the musical character of the songs in the study according to the following: complexity (1 = simple, 5 = complex), regularity (1 = random, 5 = regular), and brightness (1 = dark, 5 = bright). These three items were included as contrast to the other seven items because they should not differ in the same song between human and artificial voices. Along with the rating task, the participants were asked whether they had heard the song before, to confirm that they were unfamiliar with artificial singing voices. After a participant completed the questionnaire, it was collected, and the next stimulus was presented. The experiment took a total of approximately 20 min for each participant.

2.4. Data analysis

To extract the factor structure underlying the questionnaire items, we analysed the correlation between each pair of questionnaire items for evaluations of 24 songs averaged for the control group. We then conducted the hierarchical cluster analysis using Ward's method from a dissimilarity matrix of all pairs based on the subtraction of the absolute value of the correlation coefficient from 1 for each pair. We determined the optimal cluster number based on the comparison of the silhouette values. The analysis was carried out using Matlab 2015a (Mathworks, USA). Then, we compared the evaluations averaged across 12 songs for each voice type between the ASD and control group in each extracted factor. Two-way Analysis of Variances (ANOVAs) with a factor of group (ASD vs control) and voice type (human vs artificial) were conducted for each factor. The post hoc multiple comparison test was conducted using a *t*-test with the Bonferroni correction.

3. Results

First, we compared the impressions of the songs sung by either a human or artificial voice for each questionnaire item in each group. In line with our previous study (Tamura et al., 2015), the control participants evaluated the songs sung by the human voice more highly than those sung by the artificial voice on the