Phonetic convergence of children with and without autism in robot interaction

Yitian Hong[1], Si Chen[1,2,3,4], Fang Zhou[1], Angel Chan Tempo Tang [1,2,4]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[2]Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong SAR, China
[4]The HK PolyU-PekingU Research Centre on Chinese Linguistics, The Hong Kong Polytechnic University, Hong Kong SAR, China

Abstract

Phonetic convergence is a phenomenon in conversations where the phonetic features of interlocutors become more similar to each other. This study investigated whether children with Autism Spectrum Disorder (ASD) showed phonetic convergence of vowel formants and speech rate when interacting with a social robot. Their production was compared with their typical developing (TD) peers and the results showed that children with ASD showed convergence of vowel formants comparable to their TD peers. They also showed the capability of perceiving the change in the robot's speech rate and adjusting their own speech rate. It is suggested that the social robot may have socially motivated children with ASD, and thus they were more willing to converge their phonetic features to it compared to a human interlocutor.

Keywords: phonetic convergence, human-robot interaction, children, autism

Introduction

In Human-Human Interaction (HHI), the interlocutors coordinate their verbal cues with each other to achieve successful communication, which is referred to as phonetic convergence. During the process of coordination, they perceive the acoustic information from their conversation partners, and adjust the corresponding cues to be closer to each other. In HHI, both human interlocutors manipulate their acoustic cues in a conversation. In Human-Robot Interaction (HRI), on the other hand, the acoustic features of the robot interlocutor remain consistent without changing. The human interlocutor is the only one that might adjust the acoustic features. Therefore, HRI provides a chance for us to study phonetic convergence of human interlocutors with more controlled speech. Particularly, we carried out a study on phonetic convergence between children with Autism Spectrum Disorder (ASD) and typical developing (TD) children. Atypical social behaviour is one core feature of ASD population (American Psychiatric Association, 2013), and it has been reported

that they lack phonetic convergence in conversations with human beings (Hogstrom et al., 2018; Wynn et al., 2020). Therefore, we raised the following research questions: 1) do ASD and TD children show phonetic convergence in a conversation with a social robot? 2) do they show different convergence patterns?

Methods

Fourteen Cantonese-English bilingual children with ASD (9 male, age = 9.5±1.16) and twelve age matched TD children (8 male, age = 9.1±1.16) participated the study. Their parents signed a written consent form, and they were reimbursed for participation.

During the experiment, the children were asked to use English to interact with the social robot to find out the differences between a pair of two pictures, one is held by the participant and the other by the robot. The robot produced different scripted utterances when catching different trigger words, in normal or slow speech rate. The pictures were adapted from the pictures designed by Baker and Hazan (2011). Totally they had conversations about four pairs of pictures which are designed to elicit the same five keywords for each picture.
Their interactions were recorded. Before and after the interaction, children produced the keywords embedded in carrier sentences. The vowel formants (F1, F2) of the keywords in different conditions (before, during, and after the experiment) were extracted and compared with the corresponding keywords produced by the robot. Their speech rates were also calculated and analysed.

Results

Vowel format adjustment

The distance of F1 and F2 between the robot and children's production before (baseline), in early period (first two tasks), in late period (last two tasks) and after the interaction was calculated and fitted in the linear mixed effect model as response variable. The model revealed the main effect of time period, as shown in Table 1.

Post-hoc analysis showed that both groups reduced F1 difference in early period, compared to baseline (t=-1.68, p=0.09). But only ASD group showed a trend of further reducing the difference in late period. Both groups showed a significant reduction of difference in F2 in early (t=-3.74, p<0.001***) and late (t=-5.02, p<0.001***) periods, compared to baseline. As shown in Figure 1, when interacting with the robot, both groups of children lowered their F1 and F2 values so that their production was more similar to the robot's. However, ASD and TD children showed different manners of vowel formant convergence. TD children converged their vowel formants maximally in early

Table 1. Summary of the results from linear mixed effect models for F1 and F2 adjustment (F1 Diff = F1 of children-F1 of robot)

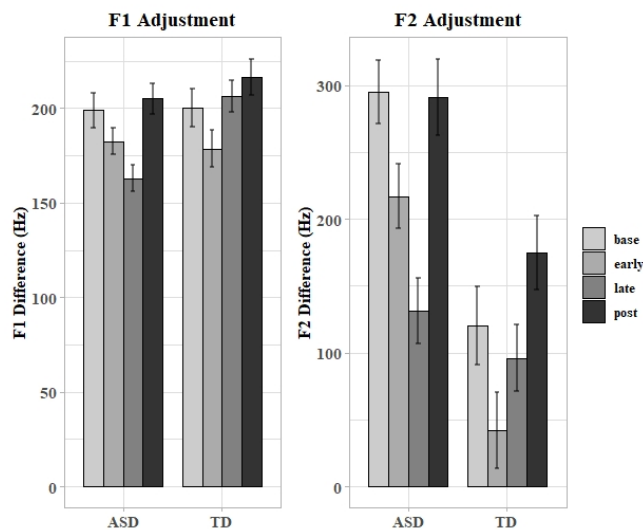| Effect	F1 Diff		F2 Diff	
	Df	p	Df	p
Period	3	.01 **	3	<0.001***
Period + Group	1	.90	1	.07
Interaction (Period * Group)	3	.12	3	.07
p <.1; * p <.05; ** p <.01; *** p <.001				

Figure 1. Mean F1 and F2 Difference between Children and Robot's Production. The error bars represent 95% confidence interval.

period and enlarged the difference in late period. In contrast, ASD children converged in a gradual manner and reached maximal convergence in the late period. It is possible that ASD children needed more time to engage themselves in the conversation and converge. It is also likely that interacting with robot for TD children was not as attractive as ASD children. They became bored in early period and were not that willing to continue the convergence in late period. Neither group showed a reduction of difference in post-interaction production, indicating that convergence of vowel formants only occurred during the interaction.

Manipulation of speech rate

Differences of two groups' children's speech rate across two conditions (slow vs. normal) were examined using a 2*2 repeated measures ANOVA (group * condition). Results showed significant effect of group (Df=1, F=8.519, p <0.01) and condition (Df=1, F=22.563, p <0.001). Their interaction did not

reach significance (Df=1, F=0.117, p =0.73), indicating that both groups adjusted their speech rate in a similar manner. Post-hoc analysis showed that TD children significantly used a faster speech rate than ASD children. Both groups reduced their speech rate significantly in slow condition, compared to normal condition, indicating their convergence to the robot's speech rate.

Discussion

In this study, we found that ASD children showed phonetic convergence of vowel formants similar to their TD peers. They were also able to perceive the manipulated speech rate of the robot and adjust their own speech rate to match it. Our findings are inconsistent with previous studies showing a lack of phonetic convergence of ASD populations. We propose two possibilities to account for this inconsistency. First, the use of social robot increased ASD children's interest of communication. Consequently, they were more engaged in HRI than HHI (previous literature) and showed more convergence. Second, as the speech of robot is consistent across the experiment, it might be easier for ASD children to detect interlocutor's phonetic features and make convergence. The controlled speech features of one interlocutor (i.e. robot) also makes the phonetic convergence of the conversation partner (i.e. child) more detectable.

Acknowledgements

References

American Psychiatric Association. 2013. DSM 5. American Journal of Psychiatry. https://doi.org/10.1176/appi.books.9780890425596.74 4053

Baker, R., Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. Behavior research methods, 43(3), 761-770.

Hogstrom, A., Theodore, R.M., Canfield, A., Castelluccio, B., Green, J., Irvine, C., Eigsti, I. M. 2018. Reduced Phonemic Convergence in Autism Spectrum Disorder. In Proceedings of the 40th Annual Conference of the Cognitive Science Society: 1797-1802, Madison, USA.

Wynn, C.J., Borrie, S.A., Sellers, T.P. 2018. Speech rate entrainment in children and adults with and without autism spectrum disorder. American Journal of Speech-Language Pathology, 27(3), 965-974.