



OPEN Multiple talker processing in autistic adult listeners

Samra Alispahic^{1✉}, Elizabeth Pellicano^{2,3}, Anne Cutler^{1,4,5} & Mark Antoniou¹

Accommodating talker variability is a complex and multi-layered cognitive process. It involves shifting attention to the vocal characteristics of the talker as well as the linguistic content of their speech. Due to an interdependence between voice and phonological processing, multi-talker environments typically incur additional processing costs compared to single-talker environments. A failure or inability to efficiently distribute attention over multiple acoustic cues in the speech signal may have detrimental language learning consequences. Yet, no studies have examined effects of multi-talker processing in populations with atypical perceptual, social and language processing for communication, including autistic people. Employing a classic word-monitoring task, we investigated effects of talker variability in Australian English autistic ($n = 24$) and non-autistic ($n = 28$) adults. Listeners responded to target words (e.g., *apple*, *duck*, *corn*) in randomised sequences of words. Half of the sequences were spoken by a single talker and the other half by multiple talkers. Results revealed that autistic participants' sensitivity scores to accurately-spotted target words did not differ to those of non-autistic participants, regardless of whether they were spoken by a single or multiple talkers. As expected, the non-autistic group showed the well-established processing cost associated with talker variability (e.g., slower response times). Remarkably, autistic listeners' response times did not differ across single- or multi-talker conditions, indicating they did not show perceptual processing costs when accommodating talker variability. The present findings have implications for theories of autistic perception and speech and language processing.

Accommodating the speech of multiple talkers is a complex and multi-layered cognitive task. It involves processing not only indexical characteristics of the talkers (e.g., differences in age, gender, emotion, social stature, dialect etc.¹), but also the content of their speech. These processes in turn affect and interact with attention and cognitive load demand². As phonetic and voice processing are highly interdependent³, challenges attending to either may have detrimental consequences for language learning, particularly for people who have altered phonetic perceptual patterns. For instance, autistic adults are able to perceive speech but exhibit relatively select voice processing atypicalities⁴. How these atypicalities interact with their ability to accommodate multiple talkers remains unclear.

It was once thought that variation between the voices of different talkers was unimportant, and that talker-specific information was discarded during the act of speech perception⁵⁻⁹. We now know, however, that talker information is not discarded because it results in a reliable processing cost during speech perception¹⁰. Many studies have demonstrated that processing the speech of multiple talkers is generally slower (and sometimes less accurate) compared to that of a single talker¹¹⁻¹⁴. Such talker variability effects occur irrespective of whether the talkers are novel or familiar to the listener¹⁵.

While several theoretical perspectives have been proposed to account for multi-talker accommodation¹⁶⁻¹⁸, the most prominent view has been termed *talker normalisation*¹⁹⁻²¹. According to this view, as the speech signal unfolds, listeners extract information about a talker's vocal characteristics^{1,22}. The listener maps the speech produced by the talker to internal phonetic categories²³. The processing required to recalibrate the vocal characteristics of multiple talkers is cognitively demanding²⁰. Therefore, the well-established processing costs of accommodating talker variability may be due to the increased cognitive demands required by talker normalisation²⁴, that is, the constant recalibrating between differing talkers' speech characteristics^{25,26}. To date, researchers have not investigated how resolving multi-talker speech affects task performance in people who may exhibit atypicalities relative to social and speech processing, namely autistic people.

¹The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia. ²Department of Educational Studies, Macquarie University, Sydney, Australia. ³Department of Clinical, Educational and Health Psychology, University College London, London, UK. ⁴Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. ⁵ARC Centre of Excellence for the Dynamics of Language, Clayton, Australia. ✉email: s.alispahic@westernsydney.edu.au

Autism affects the way a person interacts and communicates with others²⁷. While theoretical accounts have not been extended to account for speech processing, existing theoretical and empirical work offers three competing possibilities for how autistic listeners might respond to multi-talker listening environments. First, atypical voice processing has been attributed to diminished motivation to process social information in autism^{28,29}. Reduced activity in the superior temporal sulcus³⁰, a core brain region subserving processing of the human voice³¹, has been interpreted as possible “indifference” to the voices of others^{32,33}. Second, two related but distinct accounts propose a ‘detail-focused’ style of information processing, as set within Weak Central Coherence³⁴, and/or preference of lower-level processing, as described by Enhanced Perceptual Functioning³⁵. Such differences may affect higher-level language processing, resulting in superior perceptual processing of fine-grained voice information^{36,37}, but atypical speech sound specialisation or categorisation involving more complex voice stimuli, such as syllables^{38–40}.

The third possibility is that processing differences may arise from atypical predictive adaptation as derived within computational accounts (Bayesian decision theory^{41–43}), relative to processing of complex or higher-level social stimuli^{44–46}. Perceptual adaptation involves continuous (re)calibration of incoming sensory information, which match and/or update intrinsic statistics of the input⁴⁷—and these adaptive mechanisms are believed to differ in autistic people^{41,48}. Given that different talkers produce markedly different speech sounds, the listener is required to calibrate the incoming speech to the categories with the highest response probabilities⁴⁹. Such estimation of probabilities during perception are believed to differ in autistic people⁵⁰. While prediction differences have recently been proposed to be the underlying cause of language learning differences observed in autistic people, their influence on spoken language processing in autism has not been systematically investigated⁵¹.

Multi-talker speech introduces added acoustic–phonetic variability, and the resulting ambiguity affects perceptual efficiency in non-autistic listeners^{23,52–55}. However, it remains unclear how efficiently autistic listeners adapt to novel and varying voices. Autistic adult listeners’ neural processing of word-and-sentence level spoken language appears to be similar to non-autistic adults, in both clear and noisy environments^{56,57}. Likewise, processing of social characteristics of voices such as gender or age have been shown to be unaffected in autistic listeners⁵⁸. Yet, voice processing difficulties become apparent during discrimination, learning and recognition of novel compared to familiar talkers^{4,59,60}. Voice identity learning difficulties have been linked to neural differences in autistic listeners’ abilities to integrate structural lower-level voice information into a coherent percept⁶¹. If experience-dependent neural adjustment differs in autism then autistic listeners’ perceptual efficiency may also differ, relative to novel multi-talker speech. There is some evidence to suggest that autistic listeners might not show the same processing costs observed in non-autistic listeners. For example, autistic listeners are faster than non-autistic listeners when processing sung vowels by a novel talker, suggesting processing differences of vocalised sounds⁶².

Accommodating talker variability requires the allocation of additional cognitive resources (e.g., working memory), which in the general population leads to slower performance and lower accuracy in perceptual tasks⁶³. Given the limited findings concerning voice processing in autism, it is unclear how autistic people process single-versus multi-talker speech. Here, we addressed this gap by investigating multi-talker adaptation in Australian English autistic and non-autistic adults via a word-monitoring task, in which common words were spoken in two key conditions—by either (1) a single talker or (2) multiple talkers. In so doing, we tested the following three competing hypotheses:

- i. **Global Processing Difficulty:** Consistent with the common function of the superior temporal sulcus in social and speech processing, autistic adults should show a general processing difficulty, that is, prolonged response times across *both* single- and mixed-talker conditions³⁰ relative to non-autistic listeners;
- ii. **Exaggerated Cost Hypothesis:** Due to a tendency to focus on local details, thereby inducing higher levels of perceived novelty to stimuli, autistic adults should show greater processing costs (i.e., prolonged response times) when accommodating talker variability than non-autistic listeners³⁵;
- iii. **No Cost Hypothesis:** Due to differences in social functioning and diminished adaptation involving prior higher-level social stimuli, autistic adults’ responses should fall within the range of non-autistic participants, but they should not show the usual processing cost associated with talker variability⁴¹.

Methods

Participants

Australian English-born monolingual autistic ($n = 24$; 5 females) and non-autistic adults ($n = 28$; 22 females) completed this experiment. This study was conducted with the approval of Western Sydney University’s Human Research Ethics Committee. All participants provided written and verbal informed consent prior to participation, reported no hearing difficulties, and were paid a small fee (AUD\$60) for their time. Data from one additional participant (AUT group) were excluded due to an experimental script-error. The autistic and non-autistic groups were not matched for gender, $t(50) = 4.537$, $p < 0.001$, Cohen’s $d = 1.262$. Autistic participants (M age = 22.0, $SD = 8.1$, range = 18–48 years) were significantly younger than the non-autistic participants (M age = 30.3, $SD = 6.8$, range = 19–48 years), $t(51) = 3.987$, $p < 0.001$, Cohen’s $d = 1.109$. All participants were born and raised in Australia, and were monolingual speakers of English. Participants’ demographic information is presented in Table 1.

Table 2 includes participants’ age in years, and measures of intellectual functioning (Wechsler Abbreviated Scale of Intelligence—Second edition)⁶⁴, and autistic traits (Social Responsiveness Scale—Second Edition; SRS-2)⁶⁵. No between-group differences were observed in intellectual functioning across full-scale IQ, perceptual reasoning (PCI), or verbal comprehension (VCI) (all $p > 0.242$). However, as expected, SRS-2 scores were significantly higher in autistic compared to non-autistic participants, consistent with clinically-significant autistic features, $t(51) = 8.425$, $p < 0.001$, Cohen’s $d = 2.344$.

Group	Autistic	Non-autistic
<i>M</i> (SD; range) <i>N</i> (%) or frequency		
<i>N</i>	24	28
Female/male	5/19 (21/79%)	22/6 (79/21%)
Native language	Australian English	
Formal diagnosis of autism	24	0
Age of diagnosis (in years)		
0–11	19	0
12–18	2	0
18+	3	0
Other self-reported diagnoses		
Anxiety	15 (63%)	11 (39%)
Depression	8 (33%)	9 (32%)
Attention-deficit/hyperactivity disorder (ADHD)	8 (33%)	0
Dyslexia	2 (8%)	1 (4%)
Obsessive–compulsive disorder (OCD)	3 (13%)	0
Highest level of education		
School certificate	4 (17%)	0
Higher school certificate	8 (33%)	5 (18%)
Technical and further education	10 (42%)	6 (21%)
(TAFE) certificate		
Undergraduate (university)	2 (8%)	9 (32%)
Postgraduate (university)	0	8 (29%)
Currently employed		
Yes	8 (33%)	24 (86%)
No	16 (67%)	4 (14%)
Previously employed		
Yes	13 (54%)	26 (93%)
No	11 (46%)	2 (7%)
Average of hours (per week) spent interacting with people outside of immediate circle of family/friends	3.17 (SD 3.088; range = 1–15)	18.25 (SD 3.884; range = 3–20)

Table 1. Demographic information and measures of participants in the autistic and non-autistic groups.

	Group	<i>N</i>	Min	Max	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>T</i>	<i>p</i>	<i>d</i>
Age in years	AUT	24	18	58	22.04	8.084	1.65	3.987	< 0.001*	1.109
	Non-AUT	28	19	48	30.25	6.764	1.278			
SRS-2 score	AUT	24	50	79	66.88	7.903	1.613	– 8.425	< 0.001*	2.344
	Non-AUT	28	40	63	48.82	7.528	1.423			
FSIQ-4	AUT	24	94	146	117.67	15.07	3.076	– 0.557	0.580	– 0.155
	Non-AUT	28	100	131	115.75	9.474	1.79			
VCI	AUT	24	100	160	121.25	15.771	1.861	– 0.539	0.242	– 0.329
	Non-AUT	28	102	139	117	9.847	1.861			
PRI	AUT	24	86	154	112.54	17.151	2.056	– 0.593	0.556	– 0.165
	Non-AUT	28	94	133	110.21	10.881	2.056			
Interaction hours p/week	AUT	24	1	15	3.17	3.088	0.630	– 18.35	< 0.001*	4.261
	Non-AUT	28	3	20	18.25	3.884	0.734			

Table 2. Comparison of the autistic (AUT) and non-autistic (non-AUT) groups in terms of their age, SRS-2, FSIQ-4, VCI and PRI scores, and interaction hours per week. Effect size reported in Cohen's *d* with significant group differences ($p < .05$) marked with an asterisk (*). Age in years; social responsiveness scale (SRS-2) score; full scale IQ (FSIQ-4); verbal comprehension (VCI); perceptual reasoning (PCI); and hours spent interacting with people outside of their immediate circle of family and friends per week (Interaction hours p/week) were included as variables in the group statistics.

Stimuli and procedure

Stimulus materials

Stimuli in the word-monitoring task were 24 concrete nouns, that have been used in prior research^{25,66}. There were eight target words (*apple, bear, bin, cat, corn, duck, fish, and horse*), and 16 filler words (*chalk, clock, lettuce, mushrooms, notebook, onion, orange, parrot, pencil, rabbit, scissors, squirrel, stapler, strawberries, tape, and watermelon*). The words were recorded (16-bit, 44.1 kHz) from eight English monolingual talkers (4 males, 4 females). For each word, stimulus tokens were selected that had comparable durations across talkers. Tokens were root-mean-square normalised to an output level of 72 dB SPL.

Procedure

Participants were told they would be completing a word-monitoring task. At the beginning of each trial sequence, participants were presented with a target word displayed on screen in a large font. They were instructed to memorise the target word and, when ready, press the spacebar to begin. The target word would then disappear leaving a blank screen and a sequence of words would be heard through headphones. Participants were instructed to press the spacebar whenever the target word was heard. Each sequence lasted approximately 30 s, with an inter-stimulus interval of 250 ms between words. A new sequence would be indicated to the listener whereby a blank screen would flash and a new target-word would be displayed. Participants were asked to respond as quickly and accurately as possible and, if they became aware they made an error, to simply continue until the end of the experiment. This procedure closely modelled previous research^{20,25,66}.

Each sequence comprised four presentations of the target word (+ 23-filler words = 27 words per sequence). In total, sixteen sequences were presented in pseudo-randomised order: half produced by a single talker; the other by multiple talkers. Multi-talker sequences included two presentations of the target word produced by one male and one female talker, while filler words were produced by the other six talkers. Words within each sequence were randomised, but had the following constraints: target words were always separated by at least one filler; target words could not occur in the first or last position. Single- and multi-talker sequences alternated during the experiment, with single-talker sequences always being presented first.

Participants' first sequence responses were omitted from analyses to ensure that the data reflected performance only when they had become familiar with the experimental procedure (Note that we also conducted all statistical analyses with all data points included and the pattern of results was the same as those reported.). Sensitivity scores (d') were computed using the formula $z(\text{hit rate}) - z(\text{false alarm rate})$. Participants' response times for correctly identified target words were also calculated. Correct hits were defined as registered keypresses following the presentation of a target word, while false alarms were keypresses following a filler word. Responses were accepted up to 800 ms after the offset of the target word. Applying this criterion resulted in two responses being excluded (out of 3328 = 0.06% excluded).

Ethics statement

Ethics approval was obtained by the University's Human Research Ethics Committee (H3315).

Results

Autistic and non-autistic participants' word-monitoring sensitivity scores (d') are illustrated in Fig. 1.

First, a 2×2 analysis of variance (ANOVA) was conducted on the word-monitoring sensitivity scores (d'), with the between-subjects factor of group (autistic vs. non-autistic) and the within-subjects factor of talker variability (single- vs. multi-talker). The analysis revealed no significant main effects of talker variability, $F(1, 50) = 0.001$, $p = 0.975$, $\eta_p^2 = < 0.001$, or group, $F(1, 50) = 0.044$, $p = 0.834$, $\eta_p^2 = 0.001$. There was also no significant interaction between talker variability and group, $F(1, 50) = 1.37$, $p = 0.247$, $\eta_p^2 = 0.027$. Sensitivity (d') scores for correctly identified words did not differ between groups.

A second, 2×2 ANOVA was conducted on participants' response times for correctly identified words. There was a significant main effect of talker variability, $F(1, 50) = 7.96$, $p = 0.007$, $\eta_p^2 = 0.137$, reflecting faster response times in the single-talker compared to the multi-talker condition. There was no significant main effect of group, $F(1, 50) = 2.06$, $p = 0.652$, $\eta_p^2 = 0.004$, that is, autistic and non-autistic participants did not differ in their overall

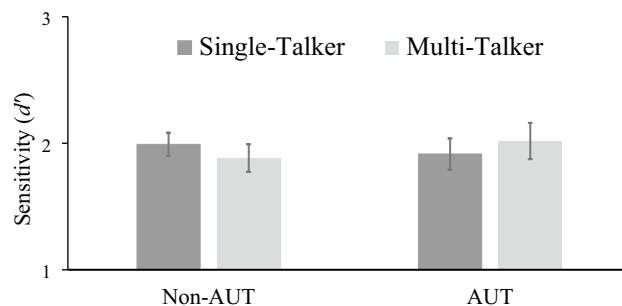


Figure 1. Non-autistic and autistic participants' sensitivity scores (d') for correctly spotted target words. Error bars represent standard error of the mean (SEM).

response times, suggesting that the autistic listeners' responses fell within the normal range. There was, however, a significant interaction between talker variability and group, $F(1, 50) = 4.59$, $p = 0.037$, $\eta_p^2 = 0.084$, suggesting that autistic and non-autistic listeners showed different patterns of word spotting speed across the single- versus multi-talker conditions (see Fig. 2).

To examine this interaction, we conducted Bonferroni-adjusted t -tests with α set to 0.025. These confirmed that the non-autistic group showed the well-established cost associated with multi-talker speech processing, $t(27) = 3.5$, $p = 0.002$, Cohen's $d = 0.664$. That is, non-autistic listeners responded faster for sequences spoken by a single talker (399 ms, $SD = 47.4$) than multiple talkers (429 ms, $SD = 52.9$). In contrast, autistic listeners *did not* show similar costs due to talker variability: for the autistic participants, there was no significant difference between response times for target words produced by a single ($M = 407$ ms, $SD = 40.6$) or by multiple talkers ($M = 411$ ms, $SD = 42.7$), $t(23) = 0.487$, $p = 0.631$, Cohen's $d = 0.099$. To further explore whether the two groups were affected differently by talker variability, we conducted an additional t -test on the difference between their response times on the single- and multi-talker conditions. This revealed a significant difference, $t(44.93) = -2.148$, $p = 0.019$, $d = -0.577$, confirming that the non-autistic group were reliably affected by talker variability (M diff = 30.3 ms, $SD = 45.7$) whereas the autistic group were not (M diff = 5.8 ms, $SD = 42.1$).

Discussion

The present study investigated the processing of multi-talker speech by autistic and non-autistic Australian-English adult listeners, testing three competing hypotheses: (i) Global Processing Difficulty; (ii) Exaggerated Cost Hypothesis; and (iii) No Cost Hypothesis. Our findings revealed that non-autistic adults showed the well-established processing costs associated with accommodating the speech of multiple talkers, namely prolonged response times relative to single-talker processing. Strikingly, the autistic group did not show this well-established pattern: for autistic listeners, response times did not differ when they processed the speech of single versus multiple talkers, despite showing accuracy levels that were indistinguishable from their non-autistic counterparts. These findings suggest an autistic advantage in processing time during multi-talker scenarios, at least as compared to non-autistic listeners.

Since autistic listeners' response times fell within the typical range of the non-autistic listeners, we found no support for the Global Processing Difficulty hypothesis. We also failed to find support for the Exaggerated Cost Hypothesis. Earlier accounts of speech perception in autistic people stated a preference for lower-level processing³⁵. From this view, if our autistic listeners were sensitive to the perceptual novelty of a talker-change, this would have prolonged response times when processing multi-talker speech. However, our autistic listeners' response times did not differ between single- and multi-talker conditions. Rather, the current results lend support to the No Cost Hypothesis, that is, differences in the use of prior higher-level social stimuli appear to confer no talker variability cost for autistic adults.

The current findings demonstrate differences between autistic and non-autistic adult listeners in the use of available speech information. Processing of complex higher-level speech was comparable in terms of sensitivity scores between listener groups; however, during spoken word recognition, talker variability led to a delay in response times in non-autistic listeners, but did not incur a similar processing cost in autistic listeners. The present findings are consistent with accounts of atypical incorporation of higher-level and lower-level processing, as set out within the Bayesian framework⁴¹. They are also in line with previous reports of attenuated perceptual mechanisms involving prior social information⁶⁷, and atypical predictive adaptation⁶⁸. Current findings further support earlier accounts of learning built through a lifetime of experience in the autistic population and atypical predictive coding integration of relevant information in spoken word processing⁶⁹.

Existing studies investigating speech processing by autistic listeners have not investigated cognitive load and response times relative to speech produced by multiple talkers^{70–74}. Cognitive demands during spoken word processing are higher when speech is produced by multiple talkers⁷⁵. A difference in rapid calibration involving higher cognitive load demands may therefore by proxy be advantageous for some autistic listeners. Therefore, and at *no cost* of accuracy, autistic individuals may be less distracted by talker variability. One explanation is that this may be evidence for an increased cognitive load capacity (i.e., working memory), as demonstrated in earlier visual inhibition tasks⁷⁶ and increased auditory perceptual capacity in detecting non-speech sounds compared to non-autistic listeners⁷⁷.

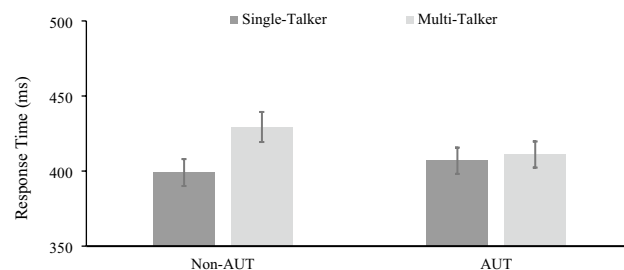


Figure 2. Autistic and non-autistic participant's response times to words produced in single and multi-talker conditions. Error bars represent standard error of the mean (SEM).

Atypical hierarchical encoding of speech may lead to differences in the speed and robustness of the early stage of auditory change detection in autistic people^{78,79}. When perceiving the speech of different talkers, listeners are required to accommodate talker variability in order to understand the linguistic utterance⁵⁵. This adds an additional processing layer that generally leads to prolonged response times, particularly when attending to more complex stimuli (e.g., words as compared to individual speech sounds)²⁰. Altered adaptation involving processes of higher-level knowledge (i.e., words), and rapid updating of lower-level linguistic categories, may affect how a new talker is detected and normalised during active listening⁸⁰. The absence of processing costs in the autistic group may therefore indicate that talker change detection differs in autistic people. Differences in talker change detection, may, in turn, affect learning outcomes relative to talker-specific speech characteristics, or when resolving ambiguity within phonetic interpretations. These results are in line with reduced adaptation⁴⁶, slower categorical updating⁸¹, and atypical phonetic encoding and language learning⁸² previously reported in autistic people.

Successful encoding of talker-related speech characteristics has been shown to have a positive impact on recognition, linguistic processing and long-term memory⁸³. The perception of a novel talker's voice interacts with the acoustic-phonetic analysis of an utterance, and talker-specific mapping is improved the more speech is heard by a listener. Talker familiarity, in turn, facilitates perceptual accuracy in the recognition of spoken words and sentences in both native¹⁷ and non-native listening⁸⁴. Therefore, the lack of talker variability effects in our autistic listeners may have emerged because they interact with fewer novel interlocutors (see Table 1) than non-autistic listeners. Processing of speech in the auditory system is facilitated by degree of social interaction⁸⁵, and increased input variability relative to social network size improves perception of speech in noise⁸⁶. Likewise, for efficient perceptual adaptation, a regular supply of novel conversational partners is required^{87,88}.

Earlier research on speech processing has questioned the impact that social factors may impose on phonological and language development in autistic children^{89,90}. Effects of smaller exposure to different talkers within a social environment have been alluded to^{45,90}, but not explicitly investigated. Autistic listeners respond faster when attending to sung vowels compared to non-vocal sounds produced by a novel talker, and compared to non-autistic listeners⁶². This does not imply that the ability to learn and recognise voices is absent in autistic people; in fact, in the same study, autistic listeners *outperformed* non-autistic listeners on voice recognition of newly encountered talkers although Lin and colleagues⁶² did not investigate how many people their autistic listeners regularly interacted with. Accounts of altered language development in autism have previously been attributed to social withdrawal⁹¹. We know, however, that autistic people are interested in social connections with others^{92–94}. Thus, investigating environmental contributions to speech perception in autism warrants further investigation.

Limitations

Due to data collection occurring during stringent COVID-19 lock-downs, the participant groups reported here were not matched in age or gender. Yet, neither gender nor age correlated with sensitivity or response times for either single- or multi-talker conditions (Supplementary Table A), suggesting limited influence of these background variables¹⁶. Further, due to not being able to access participants in-person for further testing we were also unable to test (1) whether processing patterns of multiple talkers extend to familiar (or trained) voices in autistic participants; and (2) whether a different group of autistic participants that has more in-person experience with a variety of new talkers would show the same perceptual patterns reported here. Nonetheless, the current results offer exciting possibilities for future research.

Conclusions

In sum, accommodating talker variability has been shown to exert processing costs in non-autistic listeners. However, prior work has highlighted that autistic listeners process voice and speech sounds differently to non-autistics, and this may relate to differences in flexibility during social interactions. Remarkably, our autistic listeners' performance did not differ across single- or multi-talker conditions, indicating they did not show perceptual processing costs when accommodating the speech of multiple talkers. Future research should extend this work with autistic children and adolescents, and also examine whether the effects of speech processing are relative to the amount of experience with the number of novel interlocutors, familiarity effects as well as the comparison of in-person talker processing with interactions involving other media (e.g., assistive technologies, video, television, artificial intelligence).

Data availability

The data that support the findings of this study are available in Western Sydney University Research Data open access repository via <https://doi.org/10.26183/cpk0-4b92>. Alternatively, data supporting the findings of this study are also available from the corresponding author S.A. on request.

Received: 22 November 2023; Accepted: 16 May 2024

Published online: 26 June 2024

References

- Bladon, R. A. W., Henton, C. G. & Pickering, J. B. Towards an auditory theory of speaker normalization. *Lang. Commun.* **4**, 59–69 (1984).
- Heald, S. L. M. & Nusbaum, H. C. Talker variability in audio-visual speech perception. *Front. Psychol.* **5**, 698 (2014).
- Cutler, A., Andics, A. V. & Fang, Z. Inter-dependent categorization of voices and segments. In *Proceedings of the Seventeenth International Congress of Phonetic Sciences* 552–555 (Hong Kong, 2011).
- Schelinski, S., Borowiak, K. & von Kriegstein, K. Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Soc. Cogn. Affect. Neurosci.* **11**, 1812–1822 (2016).

5. Gerstman, L. Classification of self-normalized vowels. *IEEE Trans. Audio Electroacoustics* **16**, 78–80 (1968).
6. Halle, M. Speculations about the representation of words in memory. *Phon. Linguist.* 101–114 (1985).
7. Jackson, A. & Morton, J. Facilitation of auditory word recognition. *Mem. Cognit.* **12**, 568–574 (1984).
8. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
9. Summerfield, A. & Haggard, M. P. Vocal tract normalisation as demonstrated by reaction times. in *Auditory Analysis and Perception of Speech*, (eds. G. Fant, G. & Tatham, M.A.A.) 115–141 (London, Academic Press, 1975).
10. Bradlow, A. R., Nygaard, L. C. & Pisoni, D. B. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept. Psychophys.* **61**, 206–219 (1999).
11. Mullennix, J. W. & Pisoni, D. B. Stimulus variability and processing dependencies in speech perception. *Percept. Psychophys.* **47**, 379–390 (1990).
12. Lively, S. E. & Pisoni, D. B. On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 1665–1679 (1997).
13. Francis, A. L. & Nusbaum, H. C. Paying attention to speaking rate. In *Proceeding of Fourth International Conference on Spoken Language Processing ICSLP'96*, **3**, 1537–1540 (1996).
14. Mullennix, Pisoni, D. B. & Martin, C. S. Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* **85**, 365–78 (1989).
15. Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R. & Saltzman, D. Talker familiarity and the accommodation of talker variability. *Atten. Percept. Psychophys.* **83**, 1842–1860 (2021).
16. Johnson, K. & Sjerps, M. J. Speaker normalization in speech perception. In Pardo, J. S., Nygaard, L. C., Remez, R. E. & Pisoni D. B. (Eds.), *Handb. Speech Percept.*, pp. 145–176 (Wiley-Blackwell, 2021).
17. Nygaard, L. C. & Pisoni, D. B. Talker-specific learning in speech perception. *Percept. Psychophys.* **60**, 355–376 (1998).
18. Pisoni, D. B. Some thoughts on “normalization” in speech perception. in *Talker Var. Speech Process.* (eds. Johnson, K. & Mullennix J.W.) 9–32 (Academic Press, 1997).
19. Ladefoged, P. & Broadbent, D. E. Information conveyed by vowels. *J. Acoust. Soc. Am.* **29**, 98–104 (1957).
20. Nusbaum, H. C. & Morin, T. M. Paying attention to differences among talkers. In Tohkura, Y., Vatikiotis-Bateson, E. & Sagisaka, Y. (Eds.), *Speech Percept. Prod. Linguist. Struct.*, pp. 113–134 (IOS Press, 1992).
21. Syrdal, A. K. & Gopal, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *J. Acoust. Soc. Am.* **79**, 1086–1100 (1986).
22. Nearey, T. M. Static, dynamic, and relational properties in vowel perception. *J. Acoust. Soc. Am.* **85**, 2088–2113 (1989).
23. Choi, J. Y., Hu, E. R. & Perrachione, T. K. Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Atten. Percept. Psychophys.* **80**, 784–797 (2018).
24. Magnuson, J. S. & Nusbaum, H. C. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 391–409 (2007).
25. Antoniou, M., Wong, P. C. M. & Wang, S. The effect of intensified language exposure on accommodating talker variability. *J. Speech. Lang. Hear. Res.* **58**, 722–727 (2015).
26. Antoniou, M. & Wong, P. C. M. Varying irrelevant phonetic features hinders learning of the feature being trained. *J. Acoust. Soc. Am.* **139**, 271–278 (2016).
27. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (2013).
28. Abrams, D. A. *et al.* Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *eLife* **8**, 1–33 (2019).
29. Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S. & Schultz, R. T. The social motivation theory of autism. *Trends Cogn. Sci.* **16**, 231–239 (2012).
30. Gervais, H. *et al.* Abnormal cortical voice processing in autism. *Nat. Neurosci.* **7**, 801–802 (2004).
31. Belin, P. Similarities in face and voice cerebral processing. *Vis. Cogn.* **25**, 658–665 (2017).
32. Bidet-Caulet, A. *et al.* Atypical sound discrimination in children with ASD as indicated by cortical ERPs. *J. Neurodev. Disord.* **9**, 13 (2017).
33. Klin, A. Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *J. Autism Dev. Disorders.* **21**, 29–42 (1991).
34. Happé, F. & Frith, U. The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* **36**, 5–25 (2006).
35. Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *J. Autism Dev. Disord.* **36**, 27–43 (2006).
36. Järvinen-Pasley, A., Pasley, J. & Heaton, P. Is the linguistic content of speech less salient than its perceptual features in autism?. *J. Autism Dev. Disord.* **38**, 239–248 (2008).
37. Heaton, P., Hudry, K., Ludlow, A. & Hill, E. Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cogn. Neuropsychol.* **25**, 771–782 (2008).
38. DePape, A.-M.R., Hall, G. B. C., Tillmann, B. & Trainor, L. J. Auditory processing in high-functioning adolescents with autism spectrum disorder. *PLoS ONE* **7**, e44084 (2012).
39. Lai, B. *et al.* Atypical brain lateralization for speech processing at the sublexical level in autistic children revealed by fNIRS. *Sci. Rep.* **14**, 2776 (2024).
40. Bonnel, A. *et al.* Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia* **48**, 2465–2475 (2010).
41. Pellicano, E. & Burr, D. When the world becomes ‘too real’: A Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510 (2012).
42. Sinha, P. *et al.* Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. USA.* **111**, 15220–15225 (2014).
43. van de Cruys, S. *et al.* Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
44. Cannon, J., O'Brien, A. M., Bungert, L. & Sinha, P. Prediction in Autism Spectrum Disorder: A Systematic Review of Empirical Evidence. *Autism Res. Off. J. Int. Soc. Autism Res.* **14**, 604 (2021).
45. Yu, L. *et al.* Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *J. Autism Dev. Disord.* **45**, 3656–3667 (2015).
46. Zhou, H. Y. *et al.* Audiovisual temporal processing in children and adolescents with schizophrenia and children and adolescents with autism: Evidence from simultaneity-judgment tasks and eye-tracking data. *Clin. Psychol. Sci.* **10**, 482–498 (2022).
47. Lawson, R. P., Rees, G. & Friston, K. J. An aberrant precision account of autism. *Front. Hum. Neurosci.* **8**, 302 (2014).
48. Gowen, E. & Hamilton, A. Motor abilities in autism: A review using a computational context. *J. Autism Dev. Disord.* **43**, 323–344 (2013).
49. Norris, D. & McQueen, J. M. Shortlist B: A Bayesian model of continuous speech recognition. *Psychol. Rev.* **115**, 357 (2008).
50. Gomot, M. & Wicker, B. A challenging, unpredictable world for people with autism spectrum disorder. *Int. J. Psychophysiol.* **83**, 240–247 (2012).
51. Ellis-Weismer, S. & Saffran, J. R. Differences in prediction may underlie language disorder in autism. *Front. Psychol.* **13**, 3212 (2022).

52. Choi, J. Y., Kou, R. S. N. & Perrachione, T. K. Distinct mechanisms for talker adaptation operate in parallel on different timescales. *Psychon. Bull. Rev.* **29**, 627–634 (2022).
53. Kapadia, A. M. & Perrachione, T. K. Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency. *Cognition* **204**, 104393 (2020).
54. Lim, S.-J., Carter, Y. D., Njoroge, J. M., Shinn-Cunningham, B. G. & Perrachione, T. K. Talker discontinuity disrupts attention to speech: Evidence from EEG and pupillometry. *Brain Lang.* **221**, 104996 (2021).
55. Uddin, S., Reis, K. S., Heald, S. L. M., Van Hedger, S. C. & Nusbaum, H. C. Cortical mechanisms of talker normalization in fluent sentences. *Brain Lang.* **201**, 104722 (2020).
56. Schelinski, S. & von Kriegstein, K. Responses in left inferior frontal gyrus are altered for speech-in-noise processing, but not for clear speech in autism. *Brain Behav.* **13**, e2848 (2023).
57. Callejo, D. R. & Boets, B. A systematic review on speech-in-noise perception in autism. *Neurosci. Biobehav. Rev.* 105406 (2023).
58. Groen, W. B. *et al.* Gender in voice perception in autism. *J. Autism Dev. Disord.* **38**, 1819–1826 (2008).
59. Schelinski, S., Roswadowitz, C. & von Kriegstein, K. Voice identity processing in autism spectrum disorder. *Autism Res.* **10**, 155–168 (2017).
60. Maguinness, C., Roswadowitz, C. & von Kriegstein, K. Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia* **116**, 179–193 (2018).
61. Schelinski, S. Mechanisms of voice processing: Evidence from autism spectrum disorder. *Diss.* (Humboldt University of Berlin, 2018).
62. Lin, I.-F. *et al.* Fast response to human voices in autism. *Sci. Rep.* **6**, 26336 (2016).
63. Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. Effects of stimulus variability on perception and representation of spoken words in memory. *Percept. Psychophys.* **57**, 989–1001 (1995).
64. Wechsler, D. *Wechsler abbreviated scale of intelligence—second edition (WASI-II)*. (2011).
65. Constantino, J. & Gruber, C. Social responsiveness scale: SRS-2. *Western Psychol. Serv.* (2012).
66. Sommers, M. S. & Barcroft, J. Indexical information, encoding difficulty, and second language vocabulary learning. *Appl. Psycholinguist.* **32**, 417–434 (2011).
67. Amoroso, L. *et al.* Contextual priors do not modulate action prediction in children with autism. *Proc. Biol. Sci.* **286**, (2019).
68. Turi, M., Karaminis, T., Pellicano, E. & Burr, D. No rapid audiovisual recalibration in adults on the autism spectrum. *Sci. Rep.* **6**, (2016).
69. Grisoni, L. *et al.* Prediction and mismatch negativity responses reflect impairments in action semantic processing in adults with autism spectrum disorders. *Front. Hum. Neurosci.* **13**, (2019).
70. Boucher, J. & Anns, S. Memory, learning and language in autism spectrum disorder. *Autism Dev. Lang. Impair.* **3**, 239694151774207 (2018).
71. Gilhuber, C. S., Raulston, T. J. & Galley, K. Language and communication skills in multilingual children on the autism spectrum: A systematic review. *Autism* 13623613221147780. <https://doi.org/10.1177/13623613221147780> (2023).
72. Ludlow, A. *et al.* Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *J. Autism Dev. Disord.* **8**, 697–704 (2017).
73. Matsui, T., Uchida, M., Fujino, H., Tojo, Y. & Hakarino, K. Perception of native and non-native phonemic contrasts in children with autistic spectrum disorder: Effects of speaker variability. *Clin. Ling. Phon.* **36**, 417–435 (2021).
74. Stewart, M. E., Petrou, A. M. & Ota, M. Categorical speech perception in adults with autism spectrum conditions. *J. Autism Dev. Disord.* **48**, 72–82 (2018).
75. Luthra, S., Saltzman, D., Myers, E. B. & Magnuson, J. S. Listener expectations and the perceptual accommodation of talker variability: A pre-registered replication. *Atten. Percept. Psychophys.* **83**, 2367–2376 (2021).
76. Brinkert, J. Perceptual and cognitive load in Autism—an electrophysiological and behavioural approach. *Diss.* (University College London, 2021).
77. Remington, A. & Fairnie, J. A sound advantage: Increased auditory capacity in autism. *Cognition* **166**, 459–465 (2017).
78. Font-Alaminos, M. *et al.* Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biol. Psychol.* **149**, 107807 (2020).
79. Jiang, Y. *et al.* Constructing the hierarchy of predictive auditory sequences in the marmoset brain. *eLife* **11**, e74653 (2022).
80. Heald, S. L. M. & Nusbaum, H. C. Speech perception as an active cognitive process. *Front. Syst. Neurosci.* **8**, 1–15 (2014).
81. Lieder, I. *et al.* Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat. Neurosci.* **22**, 256–264 (2019).
82. Chen, F. *et al.* Neural coding of formant-exaggerated speech and nonspeech in children with and without autism spectrum disorders. *Autism Res.* **14**, 1357–1374 (2021).
83. Theodore, R. M., Blumstein, S. E. & Luthra, S. Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Atten. Percept. Psychophys.* **77**, 1674–1684 (2015).
84. Hardison, D. M. Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Appl. Psycholinguist.* **24**, 495–522 (2003).
85. Keesom, S. M. & Hurley, L. M. Silence, solitude, and serotonin: Neural mechanisms linking hearing loss and social isolation. *Brain Sci.* **10**, 367 (2020).
86. Lev-Ari, S. The influence of social network size on speech perception. *Q. J. Exp. Psychol.* **71**, 2249–2260 (2018).
87. Bruggeman, L. & Cutler, A. No L1 privilege in talker adaptation. *Bilingualism* **23**, 681–693 (2020).
88. Burchfield, L. A., Antoniou, M. & Cutler, A. The dependence of accommodation processes on conversational experience. *Speech Commun.* **153**, 102963 (2023).
89. Kellerman, G. R., Fan, J. & Gorman, J. M. Auditory abnormalities in autism: toward functional distinctions among findings. *CNS Spectr.* **10**, 748–756 (2005).
90. Wang, X., Wang, S., Fan, Y., Huang, D. & Zhang, Y. Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Sci. Rep.* **7**, 43254 (2017).
91. Tang, B. *et al.* Evaluating causal psychological models: A study of language theories of autism using a large sample. *Front. Psychol.* **14**, 1 (2023).
92. Grace, K., Remington, A., Lloyd-Evans, B., Davies, J. & Crane, L. Loneliness in autistic adults: A systematic review. *Autism* **26**, 2117–2135 (2022).
93. Jaswal, V. K. & Akhtar, N. Being vs. appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behav. Brain Sci.* **42**, 1–84 (2018).
94. Pellicano, E. *et al.* COVID-19, social isolation and the mental health of autistic people and their families: A qualitative study. *Autism* **26**, 914–927 (2022).

Acknowledgements

This research was supported by the Australian Research Council (DP190103067). EP was supported by an Australian Research Council Future Fellowship (FT190100077). The authors would like to thank all participants and their families who very generously gave their time and patience to this study. We also thank ASPECT, the

Autism Community Network, Different Journeys, and Interchange Outer East for their invaluable support and promotion of the current project, as well as assistance in participant recruitment.

Author contributions

S.A., M.A., A.C. conceived the design of the study. S.A., E.P. and M.A. wrote the main manuscript text. S.A. collected the data. S.A. and M.A. analysed the data. S.A. prepared the Figs. 1, 2 and Tables 1, 2. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62429-w>.

Correspondence and requests for materials should be addressed to S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024