Atypical context-dependent speech processing in autism

.

Abstract: The ability to take contextual information into account is essential for successful speech processing. This study examines individuals with high-functioning autism (ASD) and those without in terms of how they adjust their perceptual expectation while discriminating speech sounds in different phonological contexts. Listeners were asked to discriminate pairs of sibilant-vowel monosyllables. Typically, discriminability of sibilants increases when the sibilants are embedded in perceptually enhancing contexts (if the appropriate context-specific perceptual adjustment were performed) and decreases in perceptually diminishing contexts. This study found a reduction in the differences in perceptual response across enhancing and diminishing contexts among ASD individuals relative to the neurotypical controls. The reduction in perceptual expectation adjustment is consistent with an increase in autonomy in low level perceptual processing in autism and a reduction in the influence of top-down information from surrounding information.}

Keywords: Context-dependent speech perception, high-functioning autism, sound discrimination, sibilant perception.

Introduction

Studies in Autism Spectrum Disorder (ASD) have identified atypicalities in auditory and speech processing among individuals with ASD. Individuals with ASD exhibit enhanced perceptual performance in auditory (and visual) domains, including not only the processing of pitch (e.g., Bonnel et al. 2003; Heaton, Davis, and Happé 2008), but show deficits in processing speech signals that involve incorporating higher order information, such as categorical perception (You et al. 2017; Stewart, Petrou, and Ota 2018) and prosody comprehension (see O'Connor 2012 for review). The atypicalities in speech processing associated with individuals with ASD might be related to differences in cognitive processing styles relative to neurotypicals, in particular relating to bias for local processing, which causes either weak top-down processing (i.e. Weak Central Coherence; Happé 1999), highly developed low-level processing (i.e. Enhanced Perceptual Functioning; Mottron et al. 2006), or a lack of flexibility in ignoring prediction errors, which leads to a focus on local processing at the expense of more abstract representations of the incoming signals (Cruys et al. 2014).

The present study investigates the effects of ASD on the discrimination of speech signals in different phonological contexts. In particular, this study focuses on how individuals with autism integrate information regarding the potential influence of one speech sound on the realization of another, commonly referred to coarticulation or coproduction. For example, perceptual studies of coarticulated speech, such as the perception of sibilants in different vocalic contexts, have found that neurotypical listeners report hearing more instances of [s] than its postalveolar counterpart, [ʃ], in the context of [u] than in the context of [a] (Mann and Repp 1980; Mitterer 2006), presumably because listeners take into account the lowered noise frequencies of /s/, which renders /s/ perceptually more [ʃ]-like, in a rounded vowel context. Previous studies on perceptual compensation for coarticulatory influence in neurotypicals have identified significant individual variability in compensatory responses. Repp (1981), for example, suggests that there exist two different strategies of listening to fricative-vowel syllables, one auditory, which segregates the noise portion from the vocalic portion, and the other phonetic, where sibilant noise information is

more integrated with the vocalic portion. More recent studies found that individual variation in perceptual compensation to be more gradient (Yu 2010; Yu and Lee 2014; Turnbull 2015). This kind of context sensitivity in speech perception or perceptual expectation adjustment strategy, commonly referred to as perceptual compensation for coarticulation, is crucial for speech comprehension since misidentifying a speech sound in a context-appropriate manner might lead to errors in phonological and lexical retrieval, which could result in miscommunication and/or sound change (Ohala 1993a, 1993b). Thus, to the extent that individuals with ASD have continuous or noncategorical perception of phonetic dimension as suggested by findings of a categorical perception deficit (You et al. 2017; Stewart, Petrou, and Ota 2018), the between-category sound difference that separates words would be less distinct. As perceptual compensation for coarticulation has been shown to be phonologically-mediated (Mitterer 2006), if individuals with ASD have difficulties discriminating category boundaries and/or integrating information from the phonological level, they are predicted to have difficulties engaging in appropriate perceptual expectation adjustments, commonly referred to as perceptual compensation or normalization. Spoken word recognition, or lexical retrieval, may then be hampered if such a listener is too focused on within-category acoustic differences at the expense of taking into account potential influence of neighboring phonological contexts on generating those differences. Thus, understanding the nature of reduced perceptual compensation for coarticulation can offer a useful window into how atypicalities in auditory or low-level phonetic perception skills interfere with the requirement of successful speech processing. That is, general difficulties with sound identification and discrimination can lead to a cascade effect on the comprehension of language at other levels, which might help explain some of the language-related problems related to ASD (Walenski, Tager-Flusberg, and Ullman 2006).

Methods

Participants

The ASD cohort consisted of fifteen Cantonese-speaking adult males with autism with ages ranging from 18 to 33. All the participants were recruited from employment programs particularly designed for young adults who have been diagnosed with high functioning ASD. The programs are run by two local non-governmental organizations in Hong Kong. ASD diagnosis was based on Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM-III) (Association, 1980) criteria and International Classification of Diseases, 10th revision (ICD-10; Organization (1990)) by either a clinical psychologist or a pediatrician during their childhood. The current state of ASD was verified by the clinical judgment of the second author who is a speech-language pathologist with ASD expertise and the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al. 2012) administered by research-reliable personnel, with a total score at or above the thresholds of autism or autism spectrum for Module 4. The hearing ability of all participants was screened with a GSI 18 screening audiometer in a sound-proofed room, with the passing criteria set at 25 dB HL at the frequencies of 1000, 2000 and 4000 Hz in both ears (American Speech-Language-Hearing Association Audiologic Assessment Panel 1996 1997). This study focuses only on male participants due to difficulty in recruiting female ASD participants in Hong Kong. All ASD participants received a nominal fee for their participation.

The neurotypical (NT) cohort included twenty male adults in Hong Kong and Chicago (N=9), all native speakers of Hong Kong Cantonese, who completed this study either for course credits or a nominal fee. Their age range was between 18 and 26 with a mean of 19.58 (SD = 1.91). None

reported any language, speech, or hearing disorders nor any mental illness. The Chicago participants were all born and raised in Hong Kong and had moved to Chicago for undergraduate or graduate education within two years prior to the time of testing.

Stimuli and Procedure

Participants performed an auditory AX discrimination task, described in Yu and Lee (2014), where the participants were asked to decide whether the consonants of two CV syllables were the same or different. To ensure maximal compatibility between earlier perceptual experiments and the current one, the stimuli were adopted from Yu and Lee (2014) (https://bit.ly/2CGcJRK); readers are referred to that study for a detailed explanation on how the stimuli were created. The target stimuli were CV syllables where the C ranges perceptually from /s/ to /ʃ/ and V is either /a/ or /u/. On each trial, two CV combinations were presented with one of two interstimulus intervals (ISI): 50ms and 750ms. The short ISI was chosen to encourage listeners to engage in the auditory model of listening which is characterized by a highly detailed but quickly decaying trace memory; with the long ISI, listeners would presumably tap into the phonological mode of listening, using a more abstracted or categorization representation of the sounds in question (Pisoni 1973). Participants were instructed to attend to the consonant and indicate whether the two consonants were different using buttons labeled SAME and DIFFERENT (button positions were counter-balanced). Participants were told the target consonants would always sound *similar* and that they should respond SAME only if they hear the targets as identical. On each trial, one target consonant was followed by /u/ and the other by /a/. The target consonants were either identical (catch trials) or differed by three steps along a 10-step series (e.g., step 1 vs. step 4, step 2 vs. step 5 etc.; discrimination trials). The effect of context was tested by comparing two conditions defined by the arrangement of the targets and the accompanying vowels in each trial. In the "enhanced" condition, target consonant with high center frequency (at [s]-end of the [s]–[ʃ] continuum) were followed by the vowel /u/ and target stimuli with low center frequency (at the [ʃ]-end of the same [s]–[ʃ] continuum) were followed by the vowel /a/ (e.g., step5/u vs. step8/a or step7/a vs. step4/u). In the "diminished" condition, the opposite arrangement is used (e.g., step8/u vs. step5/a or step4/a vs. step7/u). Based on the findings reported in Yu and Lee (2014), the discrimination of the target pairs was expected to be more accurate (i.e. easier to detect a difference between the consonants) in the "enhanced" condition than in the "diminished" condition if the listeners were engaging in perceptual compensation for coarticulation. The within-trial order of the CV pairs was counterbalanced to yield 28 unique discrimination trials and 20 unique catch trials. Finally, the natural /da/ and /du/ syllables were paired with original /s/ and /ʃ/ to create four filler pairs with an ISI of 750ms to enhance the alertness of the participant during the task. All 100 trials ((7 discrimination pairs x 2 conditions (enhanced vs. diminished)+10 catch pairs) x 2 orders (/a/-final syllable first or /u/-final syllable first) x 2 ISIs (50ms vs. 750ms) + 4 fillers) were presented in a single block and there were four repetitions of the trial block for a total of 400 trials. The order of presentation was randomized within each trial block. Participants were given a short break after two blocks.

Before the discrimination task, all participants completed a series of questionnaires online. Besides age, sex, handedness, and questions about hearing loss, speech and language disorders and mental illnesses, participants answered questions from the Autism Spectrum Quotient (AQ; Baron-Cohen et al. 2001), and two abbreviated nine-item forms of the Raven's Standard

Progressive Matrices (RSPM; Bilker et al. 2012). The AQ items were scored on a Likert scale (1-4). A total AQ score was calculated by summing all the scores for each of the items, with a maximum score of 200 and a minimum score of 50. The AQ scale was scored in such a way that a higher score is more autistic-like, i.e. lower social skills, difficulty in attention switching/strong focus of attention, high attention to detail and patterns, lower ability to communicate, and low imagination. Estimated nonverbal IQ score was assessed using the average score between the two abbreviated nine-item forms of the RSPM.

Results

The median AQ score of the NT participants was 116.5 (SD = 10.79, range = 96 – 142), compared to 102 (N = 55, SD = 14.5, range = 71–150) in Stewart and Ota (2008) and 110.05 (N = 60, SD = 18, range = 78–155) in Yu (2010). Their median RSPM score is 52.50 (SD = 3.29, range = 45.41 – 54.87). The median AQ score of the ASD participants was 132 (SD = 21, range = 89 – 168) and the median RSPM score was 52.64 (SD = 11.72, range = 17.3 – 54.86). A series of Mann-Whitney-Wilcoxon tests showed that the two cohorts do not differ in RSPM scores (Mann-Whitney $U = 141$, $p = 0.97$), while their difference in AQ scores was marginally significant (Mann-Whitney $U = 207$, $p = 0.059$). The age difference between the two cohorts is significantly different (Mann-Whitney $U = 254$, $p = 0.001$). However, comparisons between regression models with and without AGE as a predictor did not improve model-likelihood significantly; AGE was therefore not included in the following analysis.

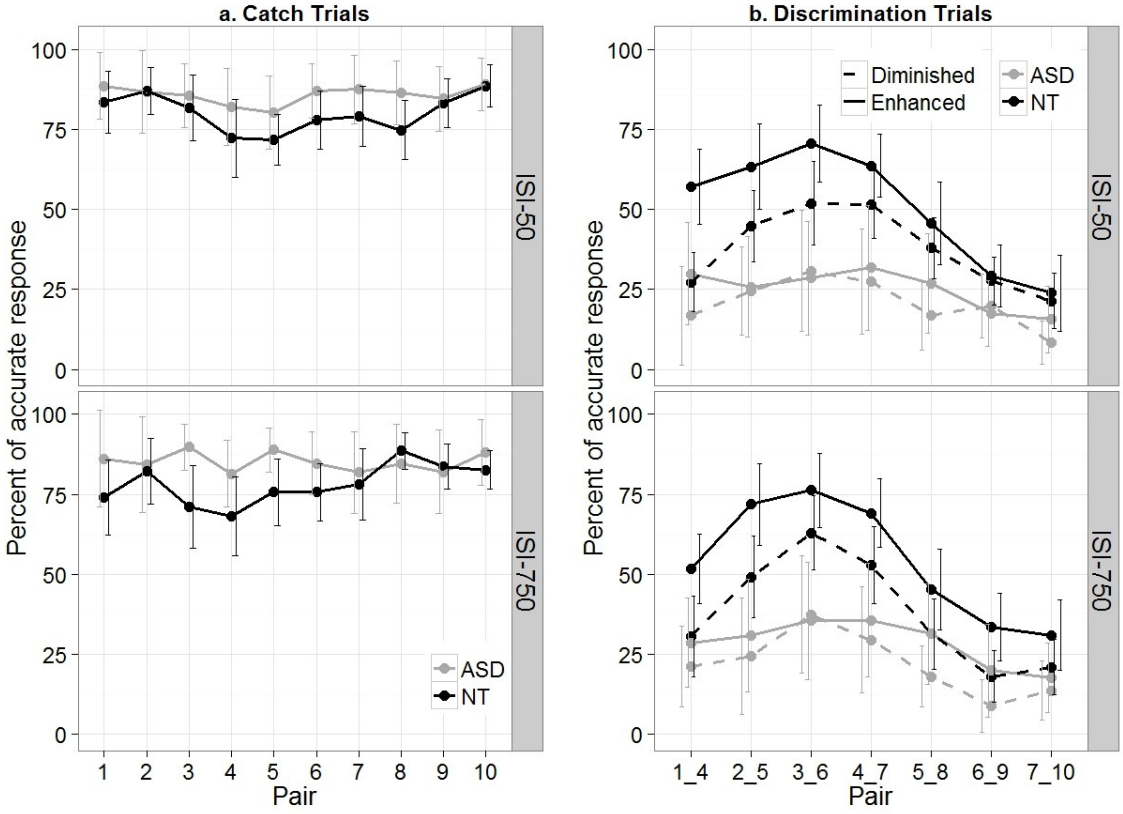

Figure 1. Mean accuracy (a) across catch pairs and (b) across "enhanced" and "diminished" pairs. The error bars indicate 95% confidence intervals.

Following Stephens and Holt (2003) and Yu and Lee (2014), response accuracy for the stimulus pairs was modeled using a series of logistic mixed-effects regressions fitted in R, using the lmer() function from the lme4 package. Responses made prior to the end of the second sibilant presentation (7% of the trials) were excluded in the regression analysis. The regression model includes the following predictors: TRIAL order (1-400), ISI (50ms vs. 750ms), CONDITION (catch vs. enhanced vs. diminished), and COHORT (ASDs vs. Neurotypicals (NT)), as well as two-way interactions between ISI and CONDITION and between CONDITION and COHORT. Neither the two-way interaction between ISI and COHORT nor the three-way interaction between ISI, CONDITION and COHORT were significant using likelihood ratio tests comparing between models with and without the particular interaction. Continuous variables were centered and scaled, and binary categorical variables were sum-coded. The CONDITION variable was Helmert-coded to allow for comparison between the catch trials and the average of the two discrimination trials (CONTRAST 1) and between the enhanced and diminished discrimination trials (CONTRAST 2). By-subject random slopes were also included for TRIAL, ISI, and CONDITION, as well as the interaction of ISI and CONDITION to allow for by-subject variability in the effect of each variable on discrimination accuracy. The model formula in lme4 style was: ACCURACY ~ TRIAL + AGE + CONDITION * (ISI + COHORT)+ (1 + TRIAL + ISI * CONDITION|SUBJECT).

Table 1. Summary of regression models for (a) response accuracy, (b) log-transformed reaction time (logRT) and (c) "different" response for all trials. For the CONDITION variable, Contrast 1 compares the catch trials with the discrimination trials while Contrast 2 compares the "enhanced" trials with the "diminished" trials. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. With respect to the linear regression model results for logRT, p-values were obtained using normal approximation which has the assumption that the *t* distribution converges to the *z* distribution as degrees of freedom increase (see Mirman 2014 for details).

	a.			b.			c.	
	Accuracy			logRT			"Different" response	
	Coef (SE)	z-value		Coef (SE)	t-value		Coef (SE)	z value
Intercept	0.008 (0.101)	0.077		6.563 (0.035)	185.239	***	-1.423 (0.248)	-5.747 ***
TRIAL	0.025 (0.034)	0.717		-0.082 (0.018)	-4.482	***	-0.063 (0.076)	-0.829
ISI	-0.031 (0.027)	-1.159		0.043 (0.009)	5.019	***	-0.067 (0.029)	-2.325 *
COHORT	-0.351 (0.096)	-3.647	***	-0.026 (0.033)	-0.771		-0.312 (0.191)	-1.636
$\text{CONDITION}_{Contrast1}$	3.231 (0.447)	7.221	***	-0.200 (0.033)	-5.994	***	-1.090 (0.176)	-6.202 ***
$\text{CONDITION}_{Contrast2}$	0.407 (0.099)	4.124	***	-0.055 (0.019)	-2.914	**	0.416 (0.101)	4.102 ***
$\text{ISI:CONDITION}_{Contrast1}$	0.120 (0.055)	2.174	*	-0.013 (0.011)	-1.159		0.062 (0.056)	1.094
$\text{ISI:CONDITION}_{Contrast2}$	-0.088 (0.054)	-1.626		-0.038 (0.018)	-2.128	*	-0.112 (0.056)	-1.992 *
$\text{COHORT:CONDITION}_{Contrast1}$	1.468 (0.441)	3.329	***	0.008 (0.032)	0.247		0.208 (0.167)	1.247
$\text{COHORT:CONDITION}_{Contrast2}$	-0.206 (0.097)	-2.127	*	-0.019 (0.019)	-1.038		-0.123 (0.097)	-1.269

A summary of the accuracy responses across conditions is given in Figure 1. Table 1a summarizes the regression model for response accuracy for all trials. There are main effects of CONDITION. Accuracy is significantly higher for the catch trials than the discrimination trials (β=3.23, z = 7.22, p< 0.001). This suggests that the participants can make relatively accurate discrimination when the sibilants are identical, but when the sibilants are different, but sufficiently close perceptually to each other, mistakes are frequent. Accuracy is significantly higher for the enhanced trials than the diminished trials (β=0.41, z = 4.12, p< 0.001). This finding indicates that, as a group, participants exhibit response patterns that are consistent with knowledge of the influence of neighboring phonological contexts on sibilant realization. There is a significant interaction between ISI and $\text{CONDITION}_{Contrast1}$, indicating that the accuracy

difference between the catch trials and the discrimination trials is larger when the ISI is short (β=0.12, z = 2.17, p< 0.05). In particular, response accuracy is higher for the catch trials and lower for the discrimination trials when listeners must rely more on auditory processing (i.e. shorter ISI).

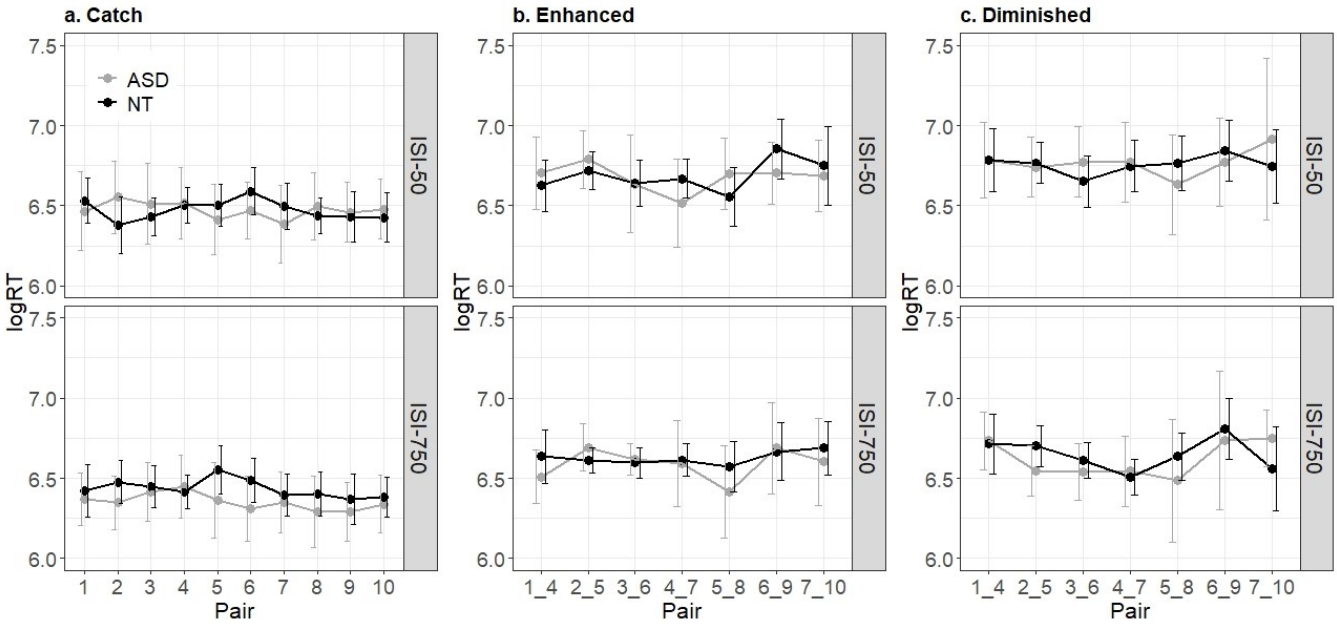

Figure 2. Mean log reaction time across cohorts and trial conditions. The error bars indicate 95% confidence intervals.

There is a main effect of COHORT (β=-0.35, z = -3.65, p< 0.001), indicating that the ASD cohort are generally less accurate compared to the NT cohort. But crucially this cohort difference is mediated by stimulus pair conditions. The model predictions for the interaction between COHORT and CONDITION are illustrated in Figure 2a. Relative to the NT cohort, the ASD cohort shows a larger accuracy difference between the catch trials and the discrimination trials (β=1.47, z = 3.33 p < 0.001). In particular, the ASD cohort is more accurate during the catch trials and less accurate during the discrimination trials. In terms of the difference between the enhanced and diminished conditions, there is also a significant difference between the ASD and NT cohorts (β=-0.21, z = -2.13, p < 0.05). While the NT cohort exhibits a sizable difference in discrimination accuracy between the enhanced and diminished conditions, the ASD cohort did not.

A summary of the (log-transformed) reaction time for the accurate trials is given in Figure 2. The (log-transformed) reaction time for the accurate trials was also analyzed in terms of a linear regression model using the same model structure as the accuracy analysis; the regression model results are summarized in Table 1b. There is a significant effect of trial order (β=-0.08, z = -4.48, p< 0.001), suggesting that participants responded faster as the experiment progressed. ISI is significant (β=0.04, z = 5.02, p< 0.001), indicating that the participants took longer to respond to the trials with short ISI than trials with long ISI. There were also significant effects of CONDITION. The participants were faster at responding to the catch trials than the discrimination trials (β=-0.20, z = -5.99, p< 0.001). Among the discrimination trials, participants were faster

with the enhanced trials than the diminished trials (β=-0.06, z = -2.91, p< 0.01). Crucially, the effects of COHORT and its interaction with CONDITION were not significant, suggesting that the NT and ASD participants patterned similarly in their processing of the stimuli, at least in terms of their reaction time relative to the different trial conditions.

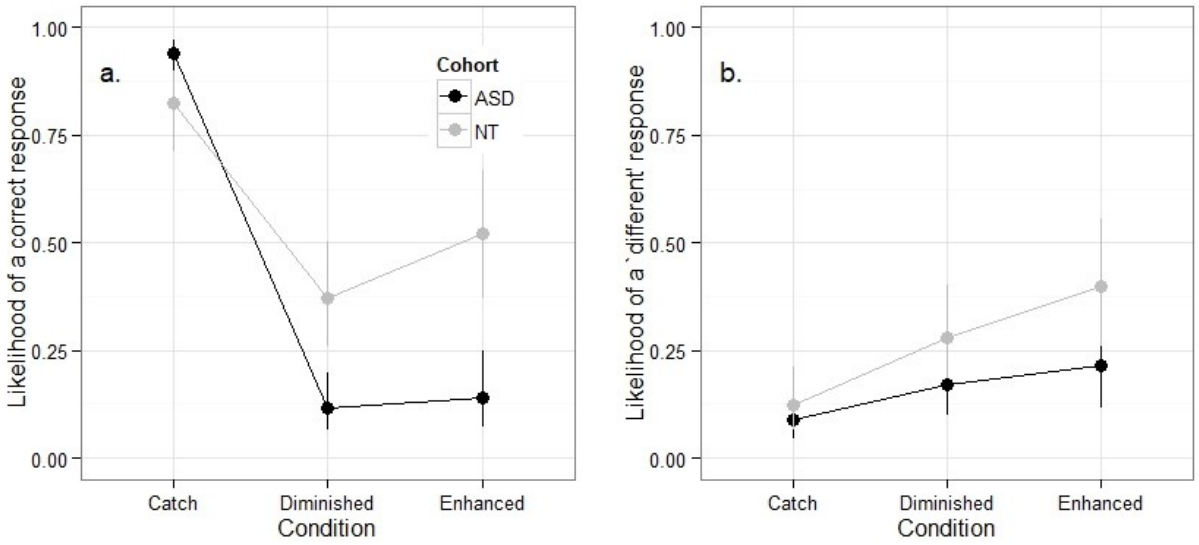

Figure 3. Model predictions for the interaction between COHORT and CONDITION in the regression models for (a) response accuracy and (b) "different" responses. The error bars indicate 95% confidence intervals.

The fact that the ASD cohort is more accurate during the catch trials and less accurate during the discrimination trials might be due to a response bias (e.g., if the ASD cohort were biased toward giving "same" responses). To examine potential biases in response patterns, a follow-up analysis on the rate of "different" response was conducted. The model structure was otherwise the same as that of the regression model for accuracy above. The regression model results are summarized in Table 1c. There was a significant effect of ISI (β=-0.07, z = -2.33, p < 0.05), suggesting that participants were less likely to respond "different" and more likely to respond "same" when the ISI was short compared to when it was long. There were also significant effects of CONDITION. As expected, there was a significant difference in "different" response between the catch trials and the discrimination trials (β=-1.09, z = -6.20, p < 0.001); a "different" response is less likely among the catch trials. The rate of "different" response is significantly higher in the enhanced condition than in the diminished one among the discrimination trials (β=0.42, z = 4.1, p < 0.001). There was also a significant interaction between ISI and $\text{Condition}_{Contrast2}$ (β=-0.11, z = -1.99, p < 0.05); the difference in "different" response rate between ISI conditions is larger in the enhanced condition than in the diminished condition. Unlike in the accuracy results, however, there was not a significant effect of COHORT nor a significant interaction between COHORT and CONDITION, suggesting that there was not a general bias toward one response by a particular cohort. As illustrated in Figure 3, the model predictions for the interaction between COHORT and CONDITION in the accuracy model (Figure 3a) and in the "different" response model (Figure 3b) show that the significant interaction between COHORT and CONDITION in the accuracy model is likely due to the larger difference in accuracy rate between the catch trials and the discrimination trials (i.e. the enhanced and diminished conditions) among the ASD cohort relative to the NT cohort. The

cohort differences in the likelihood of a "different" response across the "catch" and "discrimination" conditions are much smaller.

An examination of the accuracy rates among the discrimination trials in Figure 1 suggests that the accuracy rates differ across stimulus pairs. To further explore the differences in compensatory response between the ASD and NT cohorts, the accuracy of the responses to the discrimination trials was modeled separately in order to allow the incorporation of pair types in the analysis. It is worth noting that since a "different" response is also the accurate response in the case of the discrimination trials, an analysis of the accuracy here is also an analysis of the "different" response. In addition, the NT cohort was subdivided by the location of where the experiment took place to allow for the examination of potential language exposure effects; even though all participants have at least some familiarity with English, the NT participants in the United States who were living in an English-speaking environment might nonetheless exhibit different perceptual compensatory responses to the sibilant continua than NT participants in Hong Kong on account of the higher rate of exposure to English in Chicago. The COHORT variable was contrast-coded to allow for comparison between the ASD cohort with the NT cohort as a whole, as in the earlier models (Contrast 1) and between the ASD cohort and the NT cohort in Hong Kong in particular (Contrast 2). To reduce model complexity, the PAIR variable was reduced to three levels. That is, Level 1 consists of pairs 1_4 and 2_5 (see, e.g., the left two pairs on the x-axis in Figure 1b), while Level 2 is made up of pairs 3_6 and 4_7, where the target sibilants came from the most ambiguous region of the [s]–[ʃ] continuum. Level 3 consists of the remainder of the discrimination pairs, which are all toward the [ʃ]-end of the sibilant continuum: 5_8, 6_9, and 7_10. The PAIR variable was reverse-Helmert-coded such that the first contrast compares level 3 to the average of levels 1 & 2, while the second contrast compares between levels 1 & 2. The model structure was otherwise very similar to the regression model for accuracy above. The model formula was ACCURACY ~ TRIAL+ISI * PAIR + PAIR * CONDITION * COHORT + (1+TRIAL + ISI + PAIR + CONDITION|SUBJECT).

Table 2a summarizes the regression model for response accuracy among the discrimination trials. Besides the main effects of ISI, CONDITION, and COHORT, all replicated patterns already discussed in the first regression analysis, there was also a significant main effect of PAIR, indicating that response accuracy is higher toward the [s]-end of the continuum ($\text{PAIR}_{Contrast1}$: β=-1.07, z = -6.08, $p< 0.001$) and highest toward the middle of the continuum ($\text{PAIR}_{Contrast2}$: β=0.61, z = 3.82, $p< 0.001$). This finding is consistent with previous studies on categorical perception (e.g., Liberman et al. 1957) which show that listeners are more sensitive at discriminating across-boundary differences than within-boundary ones (i.e. the so-called "discrimination peak"). The interaction between ISI and $\text{PAIR}_{Contrast1}$ is significant ($\beta = 0.14$, z = 2.34, $p< 0.05$), indicating that the accuracy difference associated with the different ISIs is smaller in the middle than toward the /ʃ/-end of the continuum. While no significant interaction between COHORT and CONDITION was found, there were significant three-way interactions between PAIR, CONDITION, and COHORT. The model predictions for the three-way interactions are shown in Figure 4. While the heightened discrimination response in the middle of the continuum is stronger among the NT cohort than the ASD cohort, the NT cohort from Hong Kong shows a larger difference in response accuracy between the enhanced and diminished conditions ($\text{PAIR}_{contrast2}$:CONDITION:Cohort_{HK} β=0.97, z = 2.59, p= 0.01), relative to the entire NT cohort ($\text{PAIR}_{contrast2}$:CONDITION:Cohort_{all} β= -0.73, z = -2.19, p= 0.05), suggesting that the NT cohort in

Hong Kong exhibits a larger compensatory enhancement effect at the discrimination boundary relative to the ASD cohort; the enhancement effect at the discrimination boundary is smaller when the entire NT cohort is considered on account of the fact that discrimination accuracy difference between the enhanced and diminished trials are small among the Chicago-based NT cohort.

Table 2. Summary of regression models for (a) response accuracy and (b) log-transformed reaction time (logRT) of the "discrimination" trials. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. With respect to the linear regression model results for logRT, p-values were obtained using normal approximation which has the assumption that the *t* distribution converges to the *z* distribution as degrees of freedom increase.

	a.			b.		
	Accuracy			logRT		
	Coef (SE)	z-value		Coef (SE)	t-value	
Intercept	-0.767 (0.237)	-3.239	**	6.537 (0.044)	148.492	***
TRIAL	-0.002 (0.072)	-0.034		-0.081 (0.018)	-4.464	***
ISI	-0.087 (0.041)	-2.110	*	0.036 (0.007)	4.958	***
$\text{PAIR}_{Contrast1}$	-1.067 (0.176)	-6.077	***	-0.074 (0.016)	-4.514	***
$\text{PAIR}_{Contrast2}$	0.605 (0.159)	3.815	***	0.007 (0.017)	0.432	
CONDITION1	-0.295 (0.057)	-5.206	***	0.000 (0.006)	0.038	
COHORT_{All}	-2.074 (1.013)	-2.048	*	-0.008 (0.190)	-0.043	
COHORT_{HK}	0.779 (1.169)	0.666		-0.126 (0.220)	-0.572	
ISI:$\text{PAIR}_{Contrast1}$	0.136 (0.058)	2.344	*	0.011 (0.010)	1.163	
ISI:$\text{PAIR}_{Contrast2}$	-0.093 (0.074)	-1.253		0.026 (0.013)	1.991	*
$\text{PAIR}_{Contrast1}$:CONDITION1	0.118 (0.059)	1.992	*	0.003 (0.010)	0.318	
$\text{PAIR}_{Contrast2}$:CONDITION1	0.180 (0.076)	2.382	*	0.019 (0.013)	1.423	
CONDITION1:COHORT_{All}	0.241 (0.245)	0.981		-0.030 (0.026)	-1.167	
CONDITION1:COHORT_{HK}	-0.037 (0.279)	-0.133		0.044 (0.030)	1.472	
$\text{PAIR}_{Contrast1}$:COHORT_{All}	1.768 (0.749)	2.362	*	0.081 (0.074)	1.094	
$\text{PAIR}_{Contrast2}$:COHORT_{All}	0.397 (0.695)	0.571		-0.034 (0.070)	-0.485	
$\text{PAIR}_{Contrast1}$:COHORT_{HK}	-1.687 (0.854)	-1.976	*	-0.053 (0.085)	-0.619	
$\text{PAIR}_{Contrast2}$:COHORT_{HK}	-0.547 (0.791)	-0.691		0.000 (0.081)	0.006	
$\text{PAIR}_{Contrast1}$:CONDITION1:COHORT_{All}	-0.350 (0.260)	-1.346		0.001 (0.046)	0.033	
$\text{PAIR}_{Contrast2}$:CONDITION1:COHORT_{All}	-0.731 (0.334)	-2.190	*	0.003 (0.060)	0.058	
$\text{PAIR}_{Contrast1}$:CONDITION1:COHORT_{HK}	-0.051 (0.290)	-0.175		-0.013 (0.053)	-0.244	
$\text{PAIR}_{Contrast2}$:CONDITION1:COHORT_{HK}	0.970 (0.375)	2.587	**	-0.004 (0.069)	-0.053	

Table 2b summarizes the regression model for participants' log-transformed reaction time for the correct responses in the "discrimination" trials. The analysis of the log-transformed reaction time found no significant cohort related effects, suggesting that the observed accuracy differences reviewed above is not likely to be attributable simply to differences in processing patterns between cohorts.

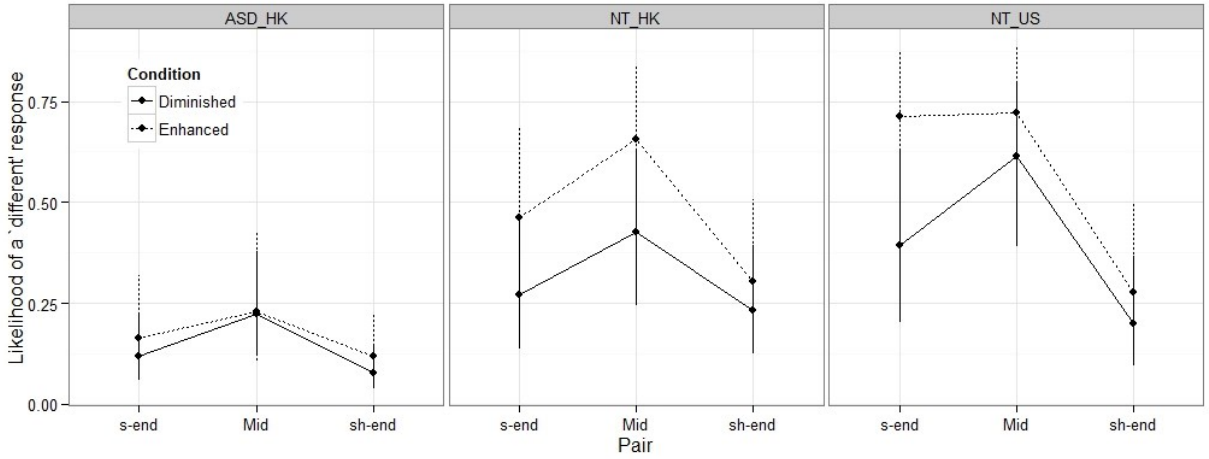

Figure 4. Model predictions for the likelihood of a "different"/accurate response across pairs in the enhanced and diminished conditions by the different cohorts. The error bars indicate 95% confidence intervals. The pairs from the [s]-end include pairs 1_4 and 2_5; the mid pairs include 3_6 and 4_7, the rest are from the [ʃ]-end.

Discussion

In general, participants tended to be more accurate in sibilant discrimination when the target sibilants were embedded in vocalic environments that maximize perceptual distinctiveness (i.e. discrimination pairs where /u/ is preceded by a target sibilant drawn from the [s]-end of the continuum while /a/ is preceded by a sibilant drawn from the [ʃ]-end of the continuum), if the listener engages in context-appropriate expectation adjustment. Accuracy rates suffered when the vocalic arrangement was in a diminished distinctiveness configuration (i.e. discrimination pairs where /a/ is preceded by a target sibilant drawn from the [s]-end of the continuum while /u/ is preceded by a sibilant drawn from the [ʃ]-end of the continuum). Crucially, the ASD and NT cohorts exhibited different compensatory response patterns. In particular, while the NT cohort showed a clear difference in discrimination accuracy across enhanced and diminished trials (i.e. perceptual compensation), the ASD cohort exhibited little differences, if at all. This reduction in the enhanced/diminished context effects in discrimination, which suggests that the ASD cohort does not benefit from the enhancement effects of the vocalic contexts, is consistent with the prediction of the "Weak Central Coherence" theory of ASD (Happé 1999), to the extent that individuals with ASD are supposed to have difficulties integrating higher order information, such as the categorical phonological identity of the neighboring sounds, in cognitive processing. These findings echo recent studies that found neurotypicals exhibiting different degrees of autistic-like traits also vary in their context-dependent speech processing behavior. Stewart and Ota (2008), for example, found that an individual's Autism-Spectrum Quotient (AQ; Baron-Cohen et al. 2001) score negatively correlates with the extent of identification shift associated with the 'Ganong effect' (i.e. the bias in categorization in the direction of a known word), after controlling for individual differences in auditory sensitivity, lexical access latency, and verbal IQ. Such a correlation is also observed in children's speech processing (Ota et al. 2015). Echoing the findings of these studies, which focus on the influence of lexical knowledge (a type of contextual information) on speech perception among individuals with different degree of autistic-like traits, our findings suggest that individuals with ASD might be less affected by

lexical knowledge in their speech perception, possibly due to their heightened sensitivity to acoustic differences and difficulties in integrating information from different levels of representation.

Our findings are also consistent with the idea that the ASD cohort is inflexible in ignoring prediction errors, focusing instead on local processing at the expense of more abstract representations of the incoming signals. In particular, given that in order to engage in context-appropriate expectation adjustment, a listener must either allow an early categorization decision to be revised in light of new information or engage in a buffered processing strategy where sound category identification is postponed until the following phonological information becomes available. From this perspective, individuals with ASD might employ a more stringent cascade processing strategy and might not be so flexible to revising their perceptual expectation once the vocalic information becomes available. The present findings are not consistent with the Enhanced Perceptual Processing theory of ASD (e.g., Mottron et al. 2006), as the theory predicts enhanced perceptual discrimination which was not seen in the current study; the response patterns observed suggest that the ASD cohort had impaired discrimination across conditions. To the extent that shorter ISI encourages phonetic (i.e. not language-specific) processing, it is noteworthy that the difference in accuracy rates across the catch and discrimination trials varies depending on the duration of the ISI. The accuracy rate is higher in the discrimination trials when the ISI is long, suggesting that discrimination enhancement effect of perceptual compensation benefits from longer processing time. This finding is consistent with the idea that perceptual compensation for coarticulation might require phonological mediation. However, the fact that an enhancement/diminished difference in accuracy is observed even when ISI is very short suggests that perceptual compensation is either not entirely dependent on phonological information or that phonological information is relevant to speech processing even at the very early stages. Our findings suggest that coarticulatory information reaches down to early perceptual processing stages (Sjerps, Mitterer, and McQueen 2011). Future neurophysiological investigations might offer better time-course information regarding the influence of coarticulatory information on speech sound representations at early stages.

As noted above, we had chosen to employ the stimuli used in Yu and Lee (2014) to ensure maximal compatibility between earlier perceptual experiments and the current one. However, this methodological choice created a potential complication for the interpretation of the results. Unlike English, which contrasts /s/ with /ʃ/, [ʃ] and [s] are allophones in Cantonese (i.e. [s] and [ʃ] do not occur in the same environment). In particular, Cantonese /s/ is more [ʃ]-like before rounded vowels and females are more likely to exhibit this allophony than males (Yu 2016). Thus, requiring native Cantonese speakers to discriminate non-native sounds from English might introduce unintended second language interference. Several factors mediate the severity of this complication. Previous studies have found that listeners are able to engage in perceptual compensation in the appropriate contexts even if the stimuli contain non-native contrasts. Mann (1986), for example, report that a group of Japanese listeners who could not identify [l] and [ɹ] accurately nonetheless showed compensation for their coarticulatory effects (see also Viswanathan, Magnuson, and Fowler 2010). Furthermore, the perceptual task was designed specifically to not require the listeners to engage in category identification of the target sounds, so the fact that [ʃ] is not a phoneme in Cantonese should have minimal impact on the participants' completion of the task. Finally, the fact that the perceptual responses of the

Cantonese-speaking NT cohort, especially the NT cohort in Hong Kong, are consistent with the findings reported in Yu and Lee (2014) suggests that the fact that the stimuli contain nonnative contrasts is not a problem for our Cantonese-speaking participants in general.

The fact that discrimination accuracy of the ASD cohort is very low overall raises questions about the possibility that the difficulty of the task itself is obscuring the ability to detect compensation for coarticulation among this population. Two factors mitigate this concern. First, as the reaction time analyses suggested, there is no significant response time difference between the ASD and NT cohorts. These findings are consistent with the interpretation that the ASD cohort was not reacting to the stimuli and the task differently from the NT cohort in terms of how much time they need to make a response decision. Second, the accuracy level of the NT-HK cohort to the stimuli in the diminished condition is not that different from that of the ASD cohort overall, suggesting that the task was difficult even for the NT cohort when the target segments were not in perceptual enhancing environments. The fact that the NT-HK cohort nonetheless shows a heightened accuracy level in the enhanced condition is consistent with the idea that the ASD cohort did not take advantage of the enhancing effects of the contextual cues.

The findings of a reduced perceptual compensatory response among the ASD cohort raise questions regarding the nature of coarticulation in speech production. Individuals with ASD are often characterized as having atypical prosody in production. Prosodic atypicalities might be related to atypicalities in degree of coarticulation in production. As the magnitude of phonological-context-dependent perceptual expectation adjustment has been shown to be positively correlated with degree of coarticulatory influence in speech production (Yu 2019), the findings of this study point to a potential reduction of coarticulation in the sibilant-vowel production among individuals with ASD. From this perspective, it is worth noting that Yu (2016) recently reported that Cantonese-speaking neurotypicals exhibiting different degree of autistic-like traits, as measured by the Autism-Spectrum Quotient (AQ; Baron-Cohen et al. 2001), also vary in the magnitude of of sibilant-vowel coarticulation they produce. In particular, individuals with higher AQ (i.e. more autistic-like traits) exhibit less sibilant-vowel coarticulation. Further studies are needed to determine whether such a reduction effect in coarticulation is observed among Cantonese-speaking individuals with ASD as well.

Acknowledgments
This research was partly supported by NSF Grants BCS-0949754 and BCS-1827409. Many thanks to the anonymous reviewers and the handling associate editor for their valuable comments and suggestions. Naturally, any errors in this work are our own.

Compliance with Ethical Standards

Conflict of Interest
The authors declare that they have no conflict of interest.

Ethical approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of Human Research Ethics Committee for Non-Clinical Faculties at the University of Hong Kong and of the Social and Behavioral Sciences Institutional Review Board

at the University of Chicago, as well as with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

American Speech-Language-Hearing Association Audiologic Assessment Panel 1996. 1997. *Guidelines for Audiologic Screening*. Rockville, MD: American Speech-Language-Hearing Association Audiologic Assessment Panel 1996.

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. "The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males, Females, Scientists and Mathematicians." *Journal of Autism & Developmental Disorders* 31: 5–17.

Bilker, Warren B., John A. Hansen, Colleen M. Brensinger, Jan Richard, Raquel E. Cur, and Ruben C. Gur. 2012. "Development of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test." *Assessment* 19: 354–69.

Bonnel, Anna, Laurent Mottron, Isabelle Peretz, Manon Trudel, Erick Gallun, and Anne-Marie Bonnel. 2003. "Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis." *Journal of Cognitive Neuroscience* 15: 226–35.

Cruys, Sander Van de, Kris Evers, Ruth Van der Hallen, Bart Boets Lien Van Eylen and, Lee de-Wit, and Johan Wagemans. 2014. "Precise Minds in Uncertain Worlds: Predictive Coding in Autism." *Psychological Review* 121 (4): 649–75.

Happé, F. 1999. "Autism: Cognitive Deficit or Cognitive Style?" *Trends in Cognitive Sciences* 3: 216–22.

Heaton, P., R. E. Davis, and F. G. E. Happé. 2008. "Research Note: Exceptional Absolute Pitch Perception for Spoken Words in an Able Adult with Autism." *Neuropsychologia* 46: 2095–8.

Liberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman, and Belver C. Griffith. 1957. "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries." *Journal of Experimental Psychology* 54: 358–68.

Lord, C., M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, and S. Bishop. 2012. *Autism Diagnostic Observation Schedule, 2nd Edition*. Torrance, CA: Western Psychological Services.

Mann, V. A. 1986. "Distinguishing Universal and Language-Dependent Levels of Speech Perception: Evidence from Japanese Listeners' Perception of English 'L' and 'R'." *Cognition* 24: 169–96.

Mann, Virginia A., and Bruno H. Repp. 1980. "Influence of Vocalic Context on Perception of the [Sh]-[S] Distinction." *Perception & Psychophysics* 28: 213–28.

Mitterer, H. 2006. "On the Causes of Compensation for Coarticulation: Evidence for Phonological Mediation." *Perception & Psychophysics* 68 (7): 1227–40.

Mottron, Laurent, Michelle Dawson, Isabelle Soulières, Benedicte Hubert, and Jake Burack. 2006. "Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception." *Journal of Autism & Developmental Disorders* 36: 27–43.

O'Connor, K. 2012. "Auditory Processing in Autism Spectrum Disorder: A Review." *Neuroscience and Biobehavioral Reviews* 36: 836–54.

Ohala, John J. 1993a. "Coarticulation and Phonology." *Language and Speech* 36 (2, 3): 155–70.

———. 1993b. “The Phonetics of Sound Change.” In *Historical Linguistics: Problems and Perspectives*, edited by C. Jones, 237–78. London: Longman Academic.
Organization, World Health. 1990. *International Classification of Diseases (10th Revision)*. Geneva: World Health Organization.
Ota, Mits, Mary Elizabeth Stewart, Alexandra Marie Petrou, and Catherine Dickie. 2015. “Lexical Effects on Children’s Speech Processing: Individual Differences Reflected in the Autism-Spectrum Quotient (AQ).” *Journal of Speech, Language, and Hearing Research* 58: 422–33.
Pisoni, David B. 1973. “Auditory and Phonetic Memory Codes in the Discrimination of Consonants and Vowels.” *Perception & Psychophysics* 13 (2): 253–60.
Repp, Bruno H. 1981. “Two Strategies in Fricative Discrimination.” *Perception & Psychophysics* 30 (3): 217–27.
Sjerps, Matthias J., Holger Mitterer, and James M. McQueen. 2011. “Listening to Different Speakers: On the Time-Course of Perceptual Compensation for Vocal-Tract Characteristics.” *Neuropsychologia* 49: 3831–46.
Stephens, Josephs D W., and Lori L. Holt. 2003. “Preceding Phonetic Context Affects Perception of Nonspeech.” *Journal of the Acoustical Society of America* 114 (6): 3036–9.
Stewart, Mary E., and Mitsuhiko Ota. 2008. “Lexical Effects on Speech Perception in Individuals with ‘Autistic’ Traits.” *Cognition* 109: 157–62.
Stewart, M. E., A. M. Petrou, and M. Ota. 2018. “Categorical Speech Perception in Adults with Autism Spectrum Conditions.” *Journal of Autism & Developmental Disorders* 48: 72–82.
Turnbull, Rory J. 2015. “Assessing the Listener-Oriented Account of Predictability-Based Phonetic Reduction.” PhD thesis, The Ohio State University.
Viswanathan, Navin, James S Magnuson, and Carol A Fowler. 2010. “Compensation for Coarticulation: Disentangling Auditory and Gestural Theories of Perception of Coarticulatory Effects in Speech.” *Journal of Experimental Psychology: Human Perception and Performance* 36 (4): 1005–15.
Walenski, Matthew, Helen Tager-Flusberg, and Michael T. Ullman. 2006. “Language in Autism.” In *Understanding Autism: From Basic Neuroscience to Treatment*, edited by S. O. Moldin and J. L. R. Rubenstein, 175–203. Boca Raton, FL: Taylor; Francis Books.
You, R. S., W. Serniclaes, D.Rider, and N. Chabane. 2017. “On the Nature of the Speech Perception Deficits in Children with Autism Spectrum Disorders.” *Research in Developmental Disabilities* 61: 158–71.
Yu, Alan Chi Lun. 2010. “Perceptual Compensation Is Correlated with Individuals’ ‘Autistic’ Traits: Implications for Models of Sound Change.” *PLoS One* 5 (8): e11950. doi:10.1371/journal.pone.0011950.
———. 2016. “Vowel-Dependent Variation in Cantonese /S/ from an Individual-Difference Perspective.” *Journal of Acoustical Society of America* 139 (4): 1672–90.
Yu, Alan C. L. 2019. “On the Nature of the Perception-Production Link: Individual Variability in English Sibilant-Vowel Coarticulation.” *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10 (1): 2.
Yu, Alan C. L., and Hyunjung Lee. 2014. “The Stability of Perceptual Compensation for Coarticulation Within and Across Individuals: A Cross-Validation Study.” *Journal of the Acoustical Society of America* 136 (1): 382–88.