ATYPICAL COARTICULATION IN AUTISM: EVIDENCE FROM SIBILANT-VOWEL INTERACTION IN CANTONESE

Alan C. L. Yu[1], Robert McAllister[1], Carol K. S. To[2]

[1]University of Chicago; [2]University of Hong Kong
aclyu@uchicago.edu; robertatlas@uchicago.edu; tokitsum@hku.hk

ABSTRACT

Atypicalities in the prosodic aspects of speech are commonly considered in clinical assessments of autism, a neurodevelopmental disorder involving difficulties in social development and communication alongside the presence of unusually strong repetitive behavior or 'obsessive' interests. While there is an increasing number of studies using objective measures to assess prosodic deficits, such studies have primarily focused on the intonational and rhythmic aspects of prosody. Little is known about prosodic deficits that are reflected at the segmental level, despite the strong connection between prosody and segmental realization. This study examines the nature of sibilant-vowel coarticulation among male adult native speakers of Cantonese with autism and those without. While neurotypical participants exhibit sibilant-vowel coarticulation that are sensitive to variation in sibilant duration, participants with autism show no sensitivity to segmental temporal changes. These findings point to the potential for atypicalities in prosody-segment interaction as an important characteristic of autistic speech.

Keywords: coarticulation, sibilant, duration, autism

1. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder involving difficulties in social development and communication alongside the presence of unusually strong repetitive behavior or 'obsessive' interests [1, 2]. Atypicalities in the prosodic aspects of speech are considered a core feature of the disorder (e.g., [3] and prosody is commonly included as a diagnostic characteristic in clinical assessments of autism [4]. Prosodic impairments are often broadly defined as deviations in speaking rate, rhythm, volume, intonation, or the inability to change language register based on the interlocutor. Individuals with ASD are often characterized with terms such as flat/monotonous, variable, sing-songy, and/or pedantic, among others (e.g., [5, 6]). While there is an increasing number of studies using objective measures to assess prosodic deficits, these studies have primarily focused on the intonational and rhythmic aspects of prosody, even though there exists strong evidence for significant interplay between articulation and different levels of linguistic structures. For example, in autosegmental metrical models of intonation, segments are organized into prosodic structures, such as syllables, feet, words, and higher phrases, as well as prominence relations in some languages (e.g., [7]). Prosodic structure is expected to influence the phonetic implementation of sound categories, and fine-grained phonetic detail, in turn, can inform higher-level prosodic structure [8]. Yet, little is known regarding prosodic deficits that may be reflected in the phonetic-prosody interface [9].

This study focuses on the nature of coarticulation in speech. Coarticulation refers to the articulatory influence between neighboring segments during speech production. The temporal extent and magnitude of coarticulation are dependent on prosodic factors of the utterance, such as speaking rate, degree of emphasis, and overall rhythm. Given that impaired temporal processing has been argued to be a source of prosodic impairments in ASD [10], we hypothesize that individuals with ASD should exhibit atypicalities in coarticulatory patterns. Specifically since temporal processing impairment not only impacts the perception of temporal variation, but also the timing accuracy in motor planning and speech production [11], we hypothesize that individuals with ASD might have difficulties with coarticulatory planning or have difficulties processing coarticulated speech [12], which in turn may lead to atypical coarticulatory patterns during production. Recent studies have found that neurotypical autistic traits are predictive of individual variation in the production and perception of coarticulated speech [13, 14], further strengthening the potential link between ASD and coarticulatory atypicalities. To investigate this link more closely, this study examines how duration variation influences the nature of sibilant-vowel coarticulation among adult male Cantonese speakers with autism and those without.

2. METHODS

2.1. Participants

The ASD cohort consisted of fifteen Cantonese-speaking adult males with autism with ages ranging from 18 to 33, with a mean of 25.51 (SD = 3.51). 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; [15]) by either a clinical psychologist or a pediatrician during their childhood. The current state of ASD was verified by the clinical judgment of the third author who is a speech-language pathologist with ASD expertise, and the Autism Diagnostic Observation Schedule (ADOS-2) [4] administered by research-reliable personnel, with a total score at or above the thresholds of autism or autism spectrum for Module 4. All particiants' hearing ability 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 [16]. This study focuses only on male participants due to difficulty in recruiting female ASD participants in Hong Kong. All ASD participation received a nominal fee for their participation.

The neurotypical (NT) cohort included twenty-three male adults in Hong Kong, all native speakers of Hong Kong Cantonese, who completed this study either for course credits or a nominal fee. Their age range was 18 to 26 with a mean of 20.13 (SD = 2.26). None reported any language, speech, or hearing disorders nor any mental illness.

2.2. Stimuli

The stimuli consisted of fourteen target words in Cantonese: [sy˥] "book,", [sy˨˩], "potato," [sɔ˥] "comb," [sɔ˨˩] "silly," [si˥] "poem," [si˨˩] "time," [se˥] "a little bit," [se˨˩] "snake," [sa˥] "sand," [san˨˩] "god," [sau˥] "to collect," [sau˥] "gloomy," [sou˥] "tassel," and [sou˨] "mob." The symbol ˥ marks a high level tone, ˨ a low tone, and ˨˩ a low falling tone.

2.3. Procedure and stimuli

Each participant was digitally recorded in a quiet room individually at a sampling rate of 44,100 Hz reading three blocks of the target stimuli, presented in traditional Chinese characters in one of two pseudo-randomized lists of target words in the carrier sentence, [ŋɔ˩˧ jiu˧ tʊk˨] ___ [pei˧˥ nei˩˧ tʰeŋ˥] "I read ___ for you to hear." A total of forty-two target stimuli were analyzed from each participant. All subjects also completed an online survey which included questions about the subject's age, sex, second language knowledge, and the full 50-question autism-spectrum quotient (AQ) questionnaire [17]. Participants were given the option to complete either the Chinese or English version of the AQ. The recordings were segmented automatically with WebMaus [18], a web service that forced-aligns an audio recording based on a corresponding orthographic transcript. German served as the language processing model as it yielded the closest alignment for these recordings. The results of the forced alignment of the target words at the segmental and word levels were manually checked and corrected where necessary.

Spectral means were extracted from syllable-initial /s/s in all target words in the corpus, using a modified version of [19], a Praat script that calculates spectral mean based on the fricative noise between the initial and final 10% of the sibilant duration high-pass filtered with a cut-off of 300 Hz using the time-averaging method [20]. Spectral mean has been shown to distinguish well between /s/ and /ʃ/ in English [21, 22] and across different vowel contexts [23, 14]. Speaking rate was calculated in terms of the number of words per second based on the average word duration between the target word and the words preceding and following the target.

3. RESULTS

The median AQ score for the ASD cohort was 132 (SD=16.27) and 121 for the NT cohort (SD=11.46). The mean sibilant duration was 197 ms and the mean speaking rate is 3.75 word per second (wps). Exploratory analyses found no significant difference between cohorts in speaking rate (ASD: mean [SD] = 3.57 [0.84] wps; NT: mean [SD] = 3.86 [0.53] wps) or sibilant duration (ASD: mean [SD] = 200.64 [36.69] ms; NT: mean [SD] = 195 [30.29] ms).

The spectral mean of /s/ was modeled using linear mixed-effects regression fitted in R, using the `lmer()` function from the `lme4` package [25]. This study focused on the effect of vocalic rounding on /s/ as it is well-documented [14]. The regression

	Model 1	Model 2
Intercept	7045.27^{***}	7092.92^{***}
	(137.46)	(136.94)
R(oun)d	-367.97^{***}	-359.95^{***}
	(58.55)	(56.87)
Cohort	-204.98	
	(135.24)	
Dur(ation)	1.10^{*}	1.39^{**}
	(0.51)	(0.53)
Rate	-165.95^{***}	-165.94^{***}
	(42.01)	(44.03)
Rd:Cohort	-38.22	
	(57.33)	
Rd:Dur	-0.80	-0.96^{*}
	(0.45)	(0.45)
Dur:Cohort	-1.05^{*}	
	(0.51)	
Rd:Dur:Cohort	1.12^{*}	
	(0.45)	
AQ		-29.35
		(135.33)
Rd:AQ		81.29
		(54.74)
Dur:AQ		-0.20
		(0.54)
Rd:Dur:AQ		0.96^{*}
		(0.46)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Table 1: Regression models of the spectral mean. The p-values were obtained using normal approximation, which assumes that the t distribution converges to the z distribution as degrees of freedom increase [24].

model tested for the effects of vocalic Rounding (rounded vs. unrounded), Cohort (ASD vs. NT), sibilant Duration, and speaking Rate. The model also included by-subject random intercepts as well as by-subject random slopes for Rate, residualized Duration, and Round, to allow for subject-specific variation with respect to these variables. Cohort was sum-coded and speaking rate centered and z-scored. Because of strong correlation between sibilant duration and speaking rate, sibilant duration was first residualized for the effect of speaking rate prior to being entered into the regression analysis. Model selection began with all interactions between Cohort, Round, and the temporal measures, i.e., Duration and Rate. Interactions that did not improve model-likelihood based on a series of likelihood ratio tests were eliminated.

The formula for the final model (Model 1 in Table 1) in `lme4` style is Spectral Mean ~ Round * Duration * Cohort + Rate + (1 + Round + Duration + Rate |Subject). As expected, there was a significant main effect of Round with spectral mean being lower before rounded vowels than before unrounded ones (β = -367.97, t-value = -6.29, $p < 0.001$). There was also a main effect of Rate (β = -165.95, t-value = -3.95, $p < 0.001$); the faster the speaking rate the lower the spectral mean. While there was a significant main effect of residualized sibilant duration (β = 1.10, t-value = 2.16, $p < 0.05$), it interacted significantly with Cohort (β = -1.05, t-value = -2.07, $p < 0.05$). As shown in Figure 1, while the ASD cohort showed a relatively stable spectral mean regardless of sibilant duration, the NT cohort exhibited a positive association between sibilant duration and spectral mean. That is, the longer the sibilant, the higher the spectral mean for the NT cohort.

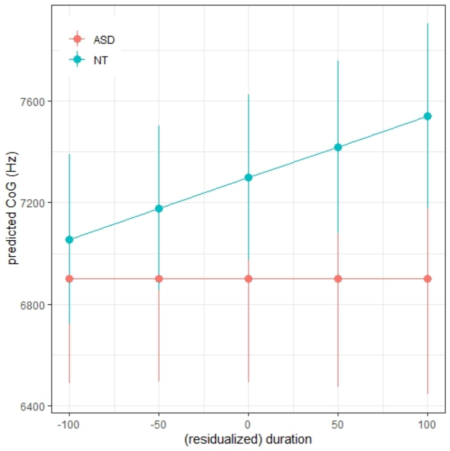

Figure 1: Effects of residualized duration sibilant spectral mean across the ASD and NT cohorts

Concerning the effect of vocalic rounding on sibilant realization, while there was not a significant interaction between Round and Cohort and between Round and Duration, there was a significant three-way interaction between Round, Duration, and Cohort (β = 1.12, t-value = 2.49, $p < 0.05$). As illustrated in Figure 2, the vocalic rounding effect on sibilant spectral mean was relatively stable for the ASD participants regardless of sibilant duration. For the NT participants, however, the vocalic rounding effect was larger the longer the sibilant. These findings suggested that, while NT speakers were sensitive to the temporal profile of the sibilant in planning the vocalic rounding influence, the ASD participants exhibited no such sensitivity. To be sure, the realization of /s/ among the ASD participants was sensitive to the roundness of the following vowel, the invariance of the context-dependent realization of /s/ suggested that the ASD participants might be treating the vocalic rounding influence on sibilant realization as a categorical phenomenon, while the NT participants treated the rounding effect on sibilant realization as more gradient and more temporally dependent. These findings seemed to be at odds with previous report concerning the effects of autistic-like traits on sibilant-vowel interaction. [14], for example, reported that Cantonese-speaking individuals with higher AQ, especially male ones, exhibit less vocalic influence in the realization of sibilants. Given that individuals with autism tend to score higher on the AQ than NT individuals, as it is the case in our sample, we might expect, *a priori*, that the ASD participants would show less vocalic influence on their sibilants than the NT participants. To be sure, [14] did not take into account of the effects of duration variation on the vocalic rounding influence on sibilant realization. Thus it is unclear how autistic-like traits might mediate such an interaction. We focused on this question in the second regression model.

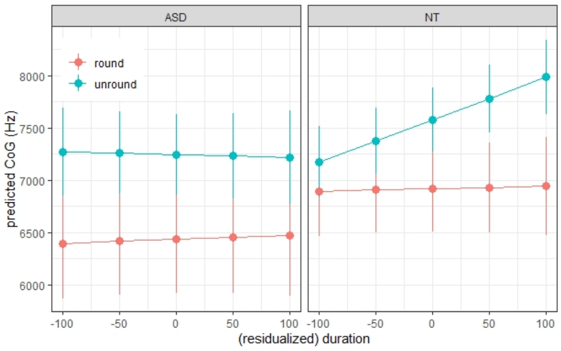

Figure 2: Three-way interaction between Round, Duration and Cohort.

The second model included all the same dependent variables except that, instead of Cohort, the participants' AQ scores were entered. Model selection proceeded the same way as the first model. The formula for the final model in `lme4` style is Spectral Mean $\sim$ Round * Duration * AQ + Rate + (1 + Round + Duration + Rate |Subject) + (1|Word). Like in the first model, there were significant main effects of Round and Duration. There was a significant interaction between Round and Duration (β = -0.96, t-value = -2.13, $p < 0.05$), indicating that the rounding effect on spectral mean increased as sibilant duration increased. While there was no main effect of AQ nor its interaction with Round or with Duration, there was a significant three-way interaction between Round, Duration, and AQ (β = 0.96, t-value = 2.08, $p < 0.05$), however. As shown in Figure 3, the higher the AQ, the smaller the effect of sibilant duration on the vocalic effect on spectral mean.

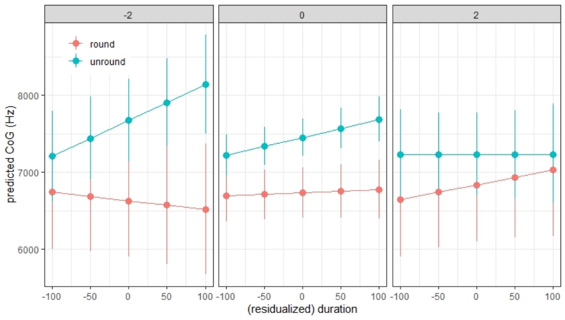

Figure 3: Three-way interaction between Round, Duration and AQ. The three panels correspond to the mean AQ (middle), and AQ scores that are two standard deviations above (right) and below (left) the mean.

4. DISCUSSION AND CONCLUSION

This study investigated the effect of sibilant duration on spectral mean as conditioned by the roundness of the following vowel among male Cantonese speakers with autism and those without. Our findings suggested that the ASD cohort exhibited no sensitivity to sibilant duration variation in producing sibilant-vowel coarticulation, while the NT cohort adjusted the magnitude of sibilant-vowel coarticulation depending on how long the sibilant was. Since sensitivity to prosodic and duration variation in realizing fine-grain phonetic difference in segmental contrast is argued to reflect speaker knowledge of controlled phonetic variation [26], the vocalic rounding effect on sibilant reflected phonetic knowledge NT speakers had. However, unlike the NT speakers, the ASD speakers showed no such temporal sensitivity, even though they exhibited spectral mean variation conditioned by vocalic rounding. These findings suggested that the ASD speakers might have internalized the vocalic influence as a categorical context-dependent allophonic variation, while the NT speakers treated the vocalic influence as gradient and is adjusted relative to the temporal dynamic of speech production. Our findings also pointed to temporal sensitivity in the realization of phonetic contrast as a potential indicator in ASD diagnosis. More research is needed to ascertain the generality of the temporal insensitivity in speech production among ASD speakers.

5. REFERENCES

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 5th ed. Washington, DC: American Psychiatric Association, 2013.

[2] I.C.D-10, Ed., *International Classification of Diseases*, 10th ed. Geneva, Switzerland: World Health Organisation, 1994.

[3] H. Tager-Flusberg, "Understanding the language and communicative impairments in autism," *International Review of Research in Mental Retardation*, vol. 23, pp. 185–205, 2000. [Online]. Available: https://doi.org/10.1016/S0074-7750(00)80011-7

[4] C. Lord, M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, and S. Bishop, *Autism Diagnostic observation schedule, 2nd edition*. Western Psychological Services, 2012.

[5] H. Amoroso, "Disorders of vocal signaling in children," in *Nonverbal vocal communication: Comparative and developmental approaches*, H. Papousek, U. Jurgens, and M. Papousek, Eds. Cambridge: Cambridge University Press, 1992, pp. 192–204.

[6] C. Lord, M. Rutter, and A. Lecouteur, "Autism Diagnostic Interview–Revised–a revisedversion of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders." *Journal of Autism and Developmental Disorders*, vol. 24, pp. 659–685, 1994.

[7] M. E. Beckman, "The parsing of prosody," *Language and Cognitive Processes*, vol. 11, pp. 17–67, 1996.

[8] T. Cho, J. McQueen, and E. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," *Journal of Phonetics*, vol. 35, pp. 210–243, 2007.

[9] T. Cho, "Laboratory phonology," in *The Continuum companion to phonology,*, N. C. Kula, B. Botma, and K. Nasukawa, Eds. London, New York: Continuum, 2011, pp. 343–368.

[10] M. Allman and C. Falter, "Abnormal timing and time perception in autism spectrum disorder?: A review of the evidence," in *Time Distortions in Mind: Temporal Processing in Clinical Populations*, A. Vatakis and M. Allman, Eds. Leiden; Boston: Brill, 2015, pp. 37–56. [Online]. Available: https://www.jstor.org/stable/https://doi.org/10.1163/j.ctt1w8h2wk.7

[11] K. Franich, H. Y. Wong, A. C. L. Yu, and C. K. S. To, "Temporal coordination and prosodic structure in autism spectrum disorder: Timing across speech and non-speech motor domains," *Journal of Autism and Developmental Disorders*, vol. 51, pp. 2929–2949, 2020. [Online]. Available: https://doi.org/10.1007/s10803-020-04758-z

[12] A. C. L. Yu and C. K. S. To, "Atypical context-dependent speech processing in autism," *Applied Psycholinguistics*, pp. 1–15, 2020.

[13] A. C. L. Yu, "Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change," *PLoS One*, vol. 5, no. 8, p. e11950, 2010. [Online]. Available: doi:10.1371/journal.pone.0011950.

[14] ——, "Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective," *Journal of Acoustical Society of America*, vol. 139, no. 4, pp. 1672–1690, 2016.

[15] W. H. Organization, *International Classification of Diseases (10th revision)*. Geneva: World Health Organization, 1990.

[16] American Speech-Language-Hearing Association Audiologic Assessment Panel 1996, *Guidelines for audiologic screening*. Rockville, MD: American Speech-Language-Hearing Association Audiologic Assessment Panel 1996, 1997.

[17] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males, females, scientists and mathematicians," *Journal of Autism & Developmental Disorders*, vol. 31, pp. 5–17, 2001.

[18] T. Kisler, U. D. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[19] C. DiCanio, "Spectral moments of fricative spectra script in praat," 2013. [Online]. Available: https://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_2.0.praat

[20] C. H. Shadle, "On the acoustics and aerodynamics of fricatives," in *The Oxford handbook of laboratory phonology*, A. C. Cohn, C. Fougeron, M. K. Huffman, and M. E. L. Renwick, Eds. Oxford: Oxford University Press, 2012, pp. 511–526.

[21] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263, 2000.

[22] C. H. Shadle and S. Mair, "Quantifying spectral characteristics of fricatives," in *ICSLP 96. Proceedings of the Fourth International Conference on Spoken Language Processing*. IEEE, 1996, pp. 1521–1524.

[23] S. Nittrouer, "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *Journal of the Acoustical Society of America*, vol. 97, pp. 520–530, 1995.

[24] D. Mirman, *Growth Curve Analysis and Visualization Using R*. Boca Raton, Florida: Chapman and Hall / CRC, 2014.

[25] D. Bates, M. Maechler, and B. Bolker, *lme4*, 2011, r package version 0.999375-38.

[26] M.-J. Solé, "Controlled and mechanical properties in speech: a review of the literature," in *Experimental Approaches to Phonology*, M.-J. Solé, P. S. Beddor, and M. Ohala, Eds. Oxford: Oxford University Press, 2007, pp. 302–321.