# EFFECTS OF WORKING MEMORY CAPACITY AND "AUTISTIC" TRAITS ON PHONOTACTIC EFFECTS IN SPEECH PERCEPTION

*Alan C. L. Yu, Julian Grove, Martina Martinović & Morgan Sonderegger*

Phonology Laboratory, University of Chicago, USA
aclyu@uchicago.edu; juliang@uchicago.edu; martinam@uchicago.edu; morgan@cs.uchicago.edu;

## ABSTRACT

Individual differences in cognitive processing style have recently been hypothesized as an important source of systematic variability in speech processing. This study offers further evidence in support of this hypothesis by showing that variability in cognitive processing style, as measured by differences in working memory capacity and "autistic" traits, significantly influences listeners' response to the effect of phonotactics in speech perception. As listeners' failure to properly normalize for context-induced variation has been taken to be a major source of innovative linguistic variants, individual variability in cognitive processing style stands to be a significant source of systematic variation in language.

**Keywords:** sound change, perceptual compensation, phonotactics, working memory, "autistic" traits

## 1. INTRODUCTION

Many researchers of sound change attribute a primary endogenous source of innovative linguistic variants to listeners failing to properly normalize for context-induced variation [4, 13]. Sources of such variation are many. In natural coarticulated speech, /s/ is acoustically more /ʃ/-like next to a rounded vowel such as /u/ or a front vowel such as /i/ due to the noise frequency lowering effect of lip protrusion or palatality respectively. Likewise, /s/ is acoustically more /ʃ/-like when produced by male talkers. Errors in perception may lead to adjustments in perceptual and production norms. Thus, in the case of sibilants, listeners might mistake a lexical item /su/ for /ʃu/ if they fail to take into account the frequency- lowering effects of the following /u/ and the listener- turned-speaker might subsequently start producing the same lexical item as [ʃu]. Repeated error of this nature could result in a drastic reduction of /s/ exemplars before /u/ and an increased number of /ʃ/ before /u/ and an *s > ʃ / u* sound change would obtain. Given that listeners are generally very good at normalizing for contextually-induced variation [8, 12], why this type of error would occur in the first place remains unclear.

Variation in cognitive processing style (CPS) — psychological dimensions representing preferences and consistencies in an individual's particular manner of cognitive functioning, with respect to acquiring and processing information — has recently been hypothesized as an important source of variation in perceptual compensation among listeners [18]. A particularly intriguing type of individual variability concerns the association between individual differences in speech perception and the extent to which individuals show "autistic" traits. Total Autism- Spectrum Quotient (AQ; [1]) [1], for individuals within the neurotypical population (i.e., those who are not diagnosed clinically as having autistic spectrum disorder), has been found to correlate significantly negatively with the extent of identification shift associated with the 'Ganong effect' (i.e., the bias in categorization in the direction of a known word) [14]. Such a finding suggests that individuals with higher degree of "autistic" traits are less likely to be affected by lexical knowledge in their phonetic perception, possibly due to their heightened sensitivity to actual acoustic differences. Likewise, while listeners generally perceive more instances of [s] than [ʃ] n the context of [u] than in the context of [a] [8] — presumably because they take into account the lowered noise frequencies of /s/ in a rounded vowel context — the magnitude of this type of perceptual compensation for vocalic coarticulation has been shown to be modulated by the listener's gender, as well as by the level of "autistic" traits s/he exhibits [18]. Individuals with low AQ, particularly women, show the smallest degree of context-specific perceptual adjustments. No significant effect of AQ was found on talker voice compensation (i.e., listeners more often identify ambiguous sibilants as /s/ when the talker is male than when the talker is female [15]).

The primary goal of the present study is to investigate how variability in CPS affects listeners' response to another important source of contextual influence in speech: phonotactics. Listeners' perceptual responses are influenced by their knowledge of what are possible and impossible sound sequences in the language. For example, when listeners were asked to classify a synthetic /r/-/l/ continuum embedded in a C_i context where C = {t, p, v, s}, they were most likely to report the ambiguous liquid as [r] when C = /t/, less likely when C = /p/ and the least when C = /v/ or /s/ [9], presumably due to the fact that *tl-* and *vr-/sr-* sequences are phonotactically ill-formed in English. Given that individual-difference dimensions such as "autistic" traits have been shown to affect perceptual compensation for coarticulation and the effects of lexical knowledge in speech perception, might differences in CPS also affect the use of phonotactic knowledge in speech perception? Two secondary effects were also tested. As noted earlier, in addition to lip rounding, the palatality of front high vowels such as [i] also exerts a frequency- lowering effect on sibilant. The present study asks to what extent perceptual compensation for the palatalization of sibilants before /i/ is also mediated by differences in CPS. Finally, to examine the possibility that perceptual compensation for talker gender is mediated by CPS differences, we present to listeners gender-neutral audio stimuli paired with faces of both genders, as it has been shown that visual cues alone are sufficient to trigger perceptual compensation for talker gender [16].

The present study focuses on two types of variability in CPS: "autistic" traits and working memory (WM) capacity. Given that the "Attention-Switching" subcomponent of the AQ has been found to have significant effects on speech perception [14, 18], we investigated whether or not variability in WM capacity affects how listeners respond to context-induced variation in speech perception, as the availability of WM resources has been shown to be positively associated with selective attention and inhibition of distracting information [5, 7].

## 2. METHODS AND MATERIALS

### 2.1. Participants

Sixty native speakers of American English (40 females) participated in the study either for course credit or for a nominal fee. All subjects performed the experiment described below, the Auto- mated Reading Span Task (RSPAN; a measure of working memory), and completed a series of on-line surveys, including the AQ.

### 2.2. Stimuli

Nine /s(C)V-ʃ(C)V/ continua were created (C = /r/ or /l/; V = /i/, /a/, or /u/). The fricative portion of the seven-step continuum was selected from a larger continuum created by digitally mixing /s/ and /ʃ/ sounds (a weighted average of the waveforms) in 2% increments. The seven fricatives were then cross-spliced with /ri/, /ra/, /ru/, /li/, /la/, /lu/, and /i/, /a/, /u/, with the final (vowel) set being taken from original /di/, /da/ and /du/. To obtain a gender ambiguous voice, the tokens recorded by a male speaker of American English were manipulated in Praat using the "Change Gender" feature and with additional manual adjustment of $f_0$. The tokens used in the experiment were judged by eight native speakers to be the most gender ambiguous among five samples of manipulated male and female voices. All tokens were normalized for pitch, duration, and intensity.

### 2.3. Procedure

Subjects were asked to identify the initial fricative as either /s/ or /ʃ/. There were two conditions with the same audio stimuli created: one with a photograph of a female face displayed on a computer screen, and the other a male face. Subjects were randomly assigned to each "face" condition. Each subject heard a total of 378 tokens (= 9 (C)V syllables x 7 steps x 3 blocks x 2 repetitions). After the identification task, subjects took the automated Reading Span task (RSPAN; [17]) to assess their WM capacity. Participants also completed the AQ.

## 3. RESULTS

### 3.1. Model

Subjects' responses (/ʃ/ vs. /s/) were modeled using mixed-effects logistic regression. The model was fitted in `R` using `lmer`, from the `lme4` package for mixed-effects models [2], with a logistic link.

**Predictors** The model contained several types of predictors. TRIAL indexed an item's order of presentation, and STEP its fricative's location on the /s(C)V-ʃ(C)V/ continuum. Three *context predictors* were included, indexing social and contextual factors expected to affect fricative perception: the following CONSONANT (/l/, /r/, none), the following VOWEL (/a/, /i/, /u/), and which FACE (male, female) was seen. Two cognitive predictors were also included: RSPAN (0-70) and AQ (50-200). Finally, SUBJECT indexes the subject associated with each item. Continuous predictors (TRIAL, STEP, RSPAN, AQ) were z-scored; CONSONANT was Helmert-coded (contrasts: none vs. l/r; l vs. r), as was VOWEL (contrasts: a vs. i/u; i vs. u); FACE was sum-coded.

*Random effects:* The model included a by-SUBJECT random intercept, to allow for subject-specific variation in /ʃ/ response rate, as well as a by-SUBJECT random slope of TRIAL, to control for subject-specific change in /ʃ/ response rate over time. Both random effects significantly improved data likelihood ($p$<0.001), when added stepwise from a model containing only fixed effects.

*Fixed effects:* A main effect term was included for each predictor (except SUBJECT). Each context predictor is expected to affect fricative perception as discussed in Section 1. Two-way interaction terms between each context predictor and each cognitive predictor were included (CONSONANT:AQ, VOWEL:RSPAN, etc.), to test whether subjects' CPS modulated the effect of each type of context. Because the slope of the identification curve has been observed to vary by context (following /a/ vs. /i/, etc.) in previous studies, the model also included two-way interaction terms between each context predictor and STEP.[2]

### 3.2. Discussion

We now summarize the model's fixed effects, omitting a full table of model results for lack of space, and discussing only terms of interest. If not discussed, a term was not significant ($p$>0.25).

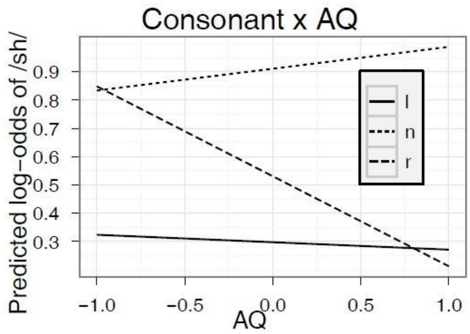
Main effects for STEP, CONSONANT, and VOWEL were as expected: the rate of /ʃ/ responses increased with increasing STEP ($p$<0.001); /ʃ/ response was lowest when the following VOWEL was /i/, higher when followed by /u/, and highest when followed by /a/ (both contrasts $p$<0.001); as for the effect of phonotactics, /ʃ/ response was highest when the onset was simplex; when the onset was complex, /ʃ/ response was lowest when the liquid was /l/ and higher when it was /r/ (both contrasts $p$<0.001). The main effect of FACE was not significant ($p$>0.3) (cf. [16]).

Both the VOWEL:RSPAN and CONSONANT:RSPAN interactions contributed significantly to data likelihood ($p$<0.001, $\chi^2(2)$=16.8; $p$<0.001, $\chi^2(2)$=46.5), while the FACE:RSPAN interaction made a nearly-marginal contribution ($p$=0.13, $\chi^2(1)$=2.3). Fig. 1 shows how the predicted effects of context predictors (VOWEL, CONSONANT, FACE) are modulated by RSPAN. It can be seen that in all three cases, the effect of context *decreases* as RSPAN increases: subjects with greater working memory capacity show less compensation for vocalic coarticulation, less influence of onset phonotactics, and perhaps less influence of perceived talker gender; however, the final pattern is not significant.

Among interactions with AQ, only the CONSONANT:AQ interaction contributed significantly to likelihood ($p$<0.001, $\chi^2(2)$=17.0); Fig. 2 shows how the predicted CONSONANT effect is modulated by AQ. The effect of phonotactics (CONSONANT=/l/ vs. /r/) is smaller for subjects with higher AQ. The VOWEL:AQ and FACE:AQ interactions did not contribute significantly to likelihood (p>0.4).

Interactions of STEP with CONSONANT and FACE contributed significantly to likelihood ($p$<0.001, $\chi^2(2)$=135.6; p<0.01, $\chi^2(1)$=7.8); while the contribution of VOWEL:STEP was marginal ($p$>0.08, $\chi^2(2)$=4.9). These interactions are not of direct interest here and will not be discussed further.

**Figure 2:** Model-predicted log-odds of /ʃ/ response as a function of AQ and CONSONANT, with other predictors held constant (as in Fig. 1). "n" here refers to syllables without a complex onset.

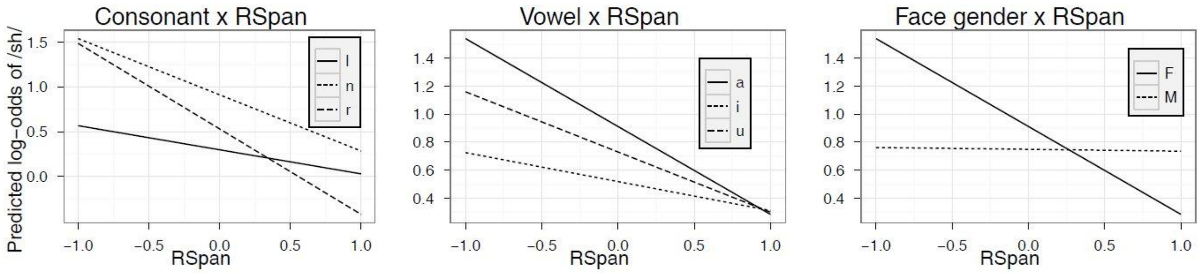



## 4. GENERAL DISCUSSION

Our findings demonstrate that differences in CPS, as measured by differences in WM and in "autistic" traits, mediate phonotactic effects in speech perception. Individuals with higher WM and a greater degree of "autistic" traits are less affected by the phonotactic context in sibilant perception. Phonotactic effects on speech perception have been attributed to the influence of the lexicon [10, 11]. Thus, the negative association between AQ and the magnitude of a phonotactic effect can be interpreted as individuals with higher AQ being less influenced by the lexicon in speech perception. In this sense, our finding is consistent with the fact that individuals with higher AQ show a weaker "Ganong effect" [14], a phenomenon certainly due to the influence of the lexicon. The strong effects of WM, independent of AQ, on all three contextual effects suggest that attentional resource is a key feature of individual differences in speech processing. The lack of an effect of AQ on vocalic perceptual compensation is surprising given that AQ was found in a previous study to be positively associated with the magnitude of perceptual compensation [18]. In light of the effect of WM capacity on perceptual compensation for vocalic coarticulation, the lack of an AQ effect on perceptual compensation for vocalic context might be due to the competing influence of lexical effects in speech perception (to which high AQ individuals are resistant) and the AQ effect on low level perceptual processing (which high AQ individuals excel in).

Further research is needed to tease apart the effects of these cognitive factors.

**Figure 1:** Model-predicted log-odds of /ʃ/ response corresponding to each significant interaction involving RSPAN. In each plot, other predictors are held constant (STEP, TRIAL, RSPAN, AQ=mean values, VOWEL=/a/, CONSO- NANT=none, FACE=F).




## 5. CONCLUSIONS

This study offers further evidence in support of the hypothesis of individual differences in CPS as a source of systematic variation in language. The present findings suggest that there exists a subsection of individuals in any speech community who regularly under-compensate and misparse. Sound change obtains to the extent that individuals with different CPS have different perceptual norms, similar differences might be reflected in the production as well, assuming that speech perception informs speech production and vice versa [3, 6].

## 6. ACKNOWLEDGEMENTS

This work is partially funded by NSF grant BCS-0949754. We thank J. Kirby for valuable comments.

## 7. REFERENCES

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

[2] Bates, D., Maechler, M., Bolker, B. 2011. *Lme4*. R package version 0.999375-38.

[3] Beddor, P.S., Harnsberger, J., Lindemann, S. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *JPhon.* 30, 591-627.

[4] Blevins, J. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

[5] Engle, R.W. 2002. Working memory capacity as executive attention. *Current Directions in Psychological Science* 11, 19-23.

[6] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *JASA* 123(5), 2825-2835.

[7] Lavie, N., Hirst, A., de Fockert, J.W., Viding, E. 2004. Load theory of selective attention and cognitive control. *JEP: General* 133, 339-354.

[8] Mann, V.A., Repp, B.H. 1980. Influence of vocalic context on perception of the [sh]-[s] distinction. *Percept Psychophys* 28, 213-228.

[9] Massaro, D.W., Cohen, M.M. 1983. Phonological constraints in speech perception. *Percept Psychophys* 34, 338-348.

[10] McClelland, J.L. 1991. Stochastic interactive processes and the effect of context on perception. *Cog. Psych*. 23, 1-44.

[11] McClelland, J.L., Elman, J.L. 1986. The TRACE model of speech perception. *Cog. Psych.* 18, 1-86.

[12] Mitterer, H. 2006. On the causes of compensation for coarticulation: Evidence for phonological mediation. *Percept Psychophys* 68(7), 1227-1240.

[13] Ohala, J. 1993. The phonetics of sound change. In Jones, C., (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman, 237-278.

[14] Stewart, M.E., Ota, M. 2008. Lexical effects on speech perception in individuals with "autistic" traits. *Cognition* 109, 157-162.

[15] Strand, E.A. 1999. Uncovering the role of gender stereotypes in speech perception. *J. Language and Psychology* 18, 86-99.

[16] Strand, E.A., Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. *Results of the 3rd KONVENS Conference* Hawthorne, 14-26.

[17] Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W. 2005. An automated version of the operation span task. *Behav Res Meth* 37(3), 498-505.

[18] Yu, A.C.L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS One* 5(8), e11950.

[1] The AQ is a self-administered scale for identifying the degree to which any individual adult of normal IQ may have traits associated with Autism Spectrum Condition.

[2] The model formula in `lme4`-style is: RESPONSE ~ (CONSONANT + VOWEL + FACE) * (AQ + RSPAN + STEP) + (1 + TRIAL | SUBJECT).