

EFFECTS OF AUTONOMIC AROUSAL AND COGNITIVE LOAD ON SENSORIMOTOR ADAPTATION OF VOICE

Authors: Nicole Tomassi¹, Alyssa Williams², Dea Turashvili³, Cara Stepp^{1,2,3,4}

Affiliations: Graduate Program for Neuroscience, Boston University¹, Department of Speech, Language, and Hearing Sciences, Boston University², Department of Biomedical Engineering, Boston University³, Department of Otolaryngology, Boston University School of Medicine⁴

Email addresses: ntomassi@bu.edu

ABSTRACT

Shared neural substrates suggest possible relationships among cognition, autonomic arousal, and speech motor control, but these systems have yet to be studied cohesively. Thus, the purpose of this study was to investigate the effects of cognitive load and autonomic arousal on sensorimotor adaptation of voice. Adults with typical speech ($n = 24$) were exposed to persistent errors to voice fundamental frequency (f_0) in two cognitive load conditions. Physiological measures of the autonomic nervous system (ANS) were simultaneously recorded to determine if changes in autonomic arousal and cognitive load were associated with the ability to adapt to these errors. Results indicated cognitive load condition as a statistically significant predictor for f_0 responses and ANS arousal changed significantly between conditions. Measures of sensorimotor adaptation, however, showed no differences between conditions.

Keywords: Sensorimotor adaptation, voice, autonomic arousal

1. INTRODUCTION

It is well known that speech production and the autonomic nervous system are related [e.g., 1, 2, 3]; however much remains unknown about the effects of the ANS on the planning and updating processes of speech, known as speech motor control. The ANS controls and regulates physiological functions involving involuntary processes. The sympathetic and parasympathetic divisions are two important branches of peripheral activity that stem from the ANS. The sympathetic branch is responsible for “fight-or-flight” responses and the parasympathetic branch for “rest-and digest” responses [4, 5]. When the ANS is aroused, the sympathetic branch is activated, and the body’s physiological functions change in response to the arousing stimuli [4, 5].

The electrodermal system is commonly targeted to measure ANS arousal because it is relatively easy to measure externally. Commercially available technology allows the estimation of electrodermal function non-invasively through wearable sensors.

The electrodermal signal measures changes in conductivity due to increases in activity of sweat glands and is a direct measure of sympathetic activation [6]. Skin conductance level and response are the most common measures derived from this signal and increase when the ANS is aroused.

Modulating cognitive load is a common way to elicit ANS arousal. ANS arousal and cognition share anatomical substrates, including the anterior cingulate cortex, as evidenced through functional imaging studies, and thus are integrated mechanistically [7-9]. Cognitive load can be modulated using various tasks employing arithmetic [e.g., 10, 11], distractor interference [e.g., 12], and executive function [e.g., 13, 14] as examples.

Error-based motor integration, or sensorimotor adaptation, is a critical component of goal-directed behavior, including for speech production. From the extensive literature aimed at understanding the neural mechanisms involved in sensorimotor adaptation, three distinct brain regions have surfaced: the anterior cingulate cortex, the cerebellum, and the basal ganglia [see review: 15]. The anterior cingulate cortex and the cerebellum are also important for autonomic regulation and cognitive function; moreover, the basal ganglia play a role in cognitive function as well. The overlapping functions among these brain regions provide theoretical reasoning as to why cognition and ANS arousal may be related to sensorimotor adaptation of speech. Yet, contemporary models of speech motor control do not account for cognitive and autonomic perspectives, leaving a gap in the understanding of the neural mechanisms involved in speech production.

Although models of speech motor control differ, most agree that neural control systems exist to ensure speech is produced accurately and efficiently [see review: 16]. Some models suggest a feedforward control system is used to send (and update when necessary) stored motor programs to speech articulators [17, 18]. Sensorimotor adaptation has been thought to be driven by the feedforward control system and can be targeted experimentally by presenting persistent and expected errors to auditory feedback [e.g., 19]. Sensorimotor adaptation is

captured by observing how the system updates and integrates motor programs to correct for the errors.

Only one prior study has described the behavioral effects of cognitive load, specifically attentional demands, on sensorimotor adaptation [20]. The study applied persistent auditory errors of voice fundamental frequency (f_0) in differing attentional demand conditions, finding significant decreases in adaptation with greater attentional demands. Although this study provided critical evidence of the impact of cognitive load on sensorimotor adaptation of voice, there are some components that need to be addressed to further fill the gap in this area. First, because ANS activity was not recorded in this study, it is unclear if the reported effects were solely due to changes in cognitive processes, or if the effects were based on physiological changes as a result of the simultaneous ANS arousal from the cognitive demands. Second, this study only used about half of the participants in the final analyses because the other half did not show adaptation in the condition without attentional load. For these reasons, the results of this study must be replicated and extended.

The purpose of the current study was to determine the contributions of cognition and autonomic arousal in error-based learning to better understand how these neural mechanisms are integrated. This question was addressed by using sensorimotor adaptation paradigms in adults with typical speech. Persistent perturbations of f_0 were compared between low and high cognitive load conditions. Electrodermal estimates of ANS activity were simultaneously collected. It was expected that the electrodermal measures would increase during the high cognitive load conditions, indicating a state of ANS arousal. We hypothesized that the updating processes of the voice would be negatively impacted by cognitive load and ANS arousal, such that adaptive responses to f_0 perturbations would decrease with increased cognitive load and increased ANS arousal.

2. METHODS

2.1. Participants

Participants were native speakers of American English (9 cisgender males, 15 cisgender females, $M_{age} = 22.76$ years, $SD_{age} = 2.66$ years) with no history of speech, language, hearing, neurological, or autonomic condition and no prior training in singing. Participants were screened to avoid potential confounds affecting ANS function or cognitive ability. Participants were non-smokers and reported no history of drug or alcohol abuse. Participants reported no prior diagnosis of mood or psychological disorders, sleep apnea, low/high blood pressure,

dermatological conditions, or hyperhidrosis. No participants were taking prescription medications known affect cognitive function. Participants abstained from the consumption of caffeine, alcohol, and heavy meals, and did not endure physical activity or stressful events three hours prior to the experiment. Participants passed a pure tone hearing screening at 25 DB HL at frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 and a color vision screening with the Ishihara Color Blindness Test [21]. All participants had typical cognitive function as assessed with the Montreal Cognitive Assessment.

2.2. Experimental design

Participants completed four speech tasks (two congruent and two incongruent) consisting of saying the color of the font of single-word stimuli ('red', 'yellow', 'black') and sustaining the midsection vowel of that color (/a/ in black or /ɜ/ in red and yellow) for approximately one second. These stimuli either matched the color of the font in all trials (congruent condition) or matched in some trials, but not in others (incongruent condition; Figure 1). During the incongruent condition, 33% of the trials were incongruent and the remaining 66% were congruent to avoid autonomic habituation. Presentation of the incongruent stimuli was semi-randomized such that incongruent trials were not sequential. In two of the four speech tasks, participants were exposed to sustained, predictable changes in auditory feedback ('shift-on') whereas in the other two tasks participants received unperturbed feedback ('control'). Thus, the four tasks were f_0 shift-on under two cognitive load conditions and control under two cognitive load conditions. The order of the tasks was counterbalanced across participants.

During the 'shift-on' tasks, f_0 was shifted

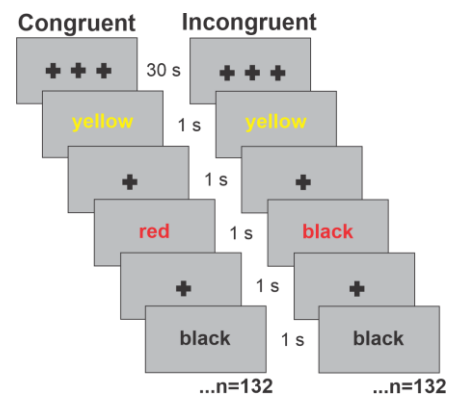


Figure 1: Single-word stimuli. Congruent: font color matches the word in all trials. Incongruent: font color matches the word in some trials, but not others.

downward over the course of four phases. The first

phase (baseline) consisted of 33 trials of unperturbed auditory feedback. The second phase, (ramp) consisted of 33 trials gradually shifted downward by 3.13 cents with each successive trial, reaching 100 cents below the participant's true f_0 . The following 33 trials in the hold phase maintained this level of feedback. The last phase (after-effect) consisted of 33 trials in which no perturbation was applied.

2.3. Instrumentation and signal processing

2.3.1. Autonomic Data

Skin conductance is sensitive to the amount of activity in eccrine sweat glands [23] and is assessed through various measures of the electrodermal signal. The electrodermal signal was conditioned using a Biopac MP150 Data Acquisition System (Biopac Systems, Inc., Goleta, CA), collected with two disposable electrodes (EL507, Biopac Systems, Inc., Goleta, CA) and an amplifier (GSR 100C, Biopac Systems, Inc., Goleta, CA), and sampled at 14700 Hz using a National Instrument data acquisition card. Electrodes were attached to the palmar surface of the first and second distal phalange on the right hand. The tonic skin conductance level was amplified with a gain of 10 $\mu\text{S}/\text{V}$ via a constant voltage (0.5 V). The phasic skin conductance response was derived from the tonic signal using a second-order Chebyshev high-pass filter with a cut-off frequency of 0.07 Hz.

2.3.2. Adaptation Data

Participants wore a SHURE MX153 omnidirectional microphone positioned 45 degrees from the midline and 7 cm from the lips and Sennheiser HD 280 Pro over-ear headphones. The microphone signal was pre-amplified using a RME Quadmic II and sampled using a RME Fireface UCX at a rate of 44,100 Hz and 32-bit resolution. Auditory feedback was calibrated to +5 dB relative to the microphone [24]. The Eventide Eclipse V4 Harmonizer was used to manipulate f_0 .

2.4. Data analysis

2.4.1. Autonomic data analysis

Electrodermal activity was derived from the skin conductance signal and calculated using a custom MATLAB graphical user interface. *Skin conductance response (SCR) count* was used to capture the number of phasic responses. SCR count was calculated for each phase of the task and used in the regression analysis in determining its effects on f_0 . Due to high variability in ANS data across participants, the change in SCR count between conditions was

calculated for each participant and the average of these changes was used to assess group-level arousal.

2.4.2. Adaptation data analysis

Estimation of f_0 was performed offline using an autocorrelation method via Praat scripts [25]. Analysis of adaptive responses was performed offline using custom MATLAB scripts. The mean f_0 of every trial in each condition was calculated over a time window of 40-120 ms [26] after vowel onset, to avoid coarticulation effects and feedback-based adjustments [27-29], and converted to cents relative to the mean f_0 during the baseline phase. The control conditions were subtracted from the shift conditions to account for prosody and natural variability over time. Adaptive responses were calculated as the mean f_0 during the hold phase after the normalization processes. The average f_0 of each phase for all participants was used in regression analyses.

2.5. Statistical analysis

All statistical analyses were performed in Minitab 21.2. To account for group-level variability in ANS arousal, changes were determined using the average between-condition change in SCR count. One-sample t-tests were used to determine whether between-condition changes in ANS arousal were greater than zero. Overall adaptation was assessed in each condition using the average f_0 responses during the hold and after-effect phases. One sample t-tests were used to determine if f_0 responses during these phases were greater than zero, indicating adaptation. Differences in adaptation between conditions were assessed using two-sample t-tests. A regression analysis was performed to comprehensively assess the effect of cognitive load (congruent vs. incongruent), phase (baseline, ramp, hold, and after-effect), and ANS arousal (SCR count) on f_0 responses.

3. RESULTS

The one-sample t-tests for individual changes in SCR count between cognitive load conditions revealed a statistically significant change. Specifically, the average change in SCR count between conditions was 6.00 peaks (SD = 4.81 peaks); $t(23) = 5.72, p < .001$. Group results for f_0 responses are illustrated in Figure 2. One-sample t-tests showed significant f_0 responses greater than zero in the hold phase for both conditions: congruent $t(23) = 2.54, p = .010$, incongruent $t(23) = 3.79, p = .001$ and in the after-effect phase for the incongruent condition: $t(23) = 1.88, p = .037$. One-sample t-tests did not show significant f_0 responses greater than zero in the after-effect phase in the congruent condition: $t(23) = -0.13$,

$p = .550$ The two-sample t-tests for the changes in adaptation between conditions did not reveal a significant difference in either phase: hold $t(46) = -0.81, p = .431$, after-effect $t(46) = -1.46, p = .076$.

Predictor Variables	df	F	p
SCR Count	1	3.58	.063
Phase	3	7.66	<.001*
Condition	1	11.05	.001*
Condition × Phase	3	1.46	ns

Table 1: Results of the regression analyses. * Indicates statistically significant (p-value < .05) predictor variables.

The results of the regression model are described in Table 1. Condition was a significant predictor of f_0 responses, $F(1, 23) = 11.05, p = .001$ with larger responses in the incongruent condition. Phase was found to be a significant predictor in the model, $F(3, 23) = 7.66, p < .001$ with post-hoc Tukey tests confirming a significantly larger mean f_0 during the hold phase, $p < .05$. SCR count was not a significant predictor of f_0 responses, $F(1,23) = 3.58, p = .063$, nor was the condition and phase interaction, $F(3,23) = 1.46, p = .477$.

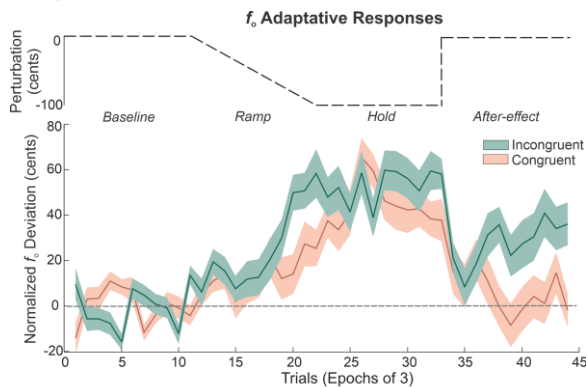


Figure 2: Group f_0 responses to altered auditory feedback.

4. DISCUSSION

Our results indicate an increase in ANS arousal between conditions. This shows that the cognitive load task implemented in the current study successfully aroused the ANS. Although SCR count was not a significant predictor of changes in f_0 , further investigation is warranted within other systems that rely on sympathetic and parasympathetic activity, such as the cardiovascular system. The f_0 responses were significantly greater than zero in the hold phase for both conditions, indicating that participants adapted to the change in auditory feedback; further, the responses were not significantly different between conditions in either phase. Thus, the incongruent load did not diminish adaptation, as predicted, but perhaps,

elicited longer-lasting adaptation, relative to the congruent condition, with f_0 responses being significantly greater than zero in the after-effect phase. Interestingly, condition was a significant predictor of f_0 responses in the regression analysis, likely due to the power of including responses in all phases, rather than the average of only the hold and after-effect phases as with the t-tests.

This result is interesting because it differs from a prior study finding [20]. These findings may be different from the findings of Scheerer, et al. [20] for a number of reasons. First, the current study used data from all participants, regardless of behavior to altered auditory feedback, whereas Scheerer, et al. [20] excluded 14 out of 30 participants for not compensating in the low cognitive load condition. A second reason as to why these results may differ from Scheerer, et al. [20], could be the differences in the cognitive tasks employed. Scheerer, et al. [20] used an attention task, whereas the current study used a response inhibition task. Although attention and response inhibition require skills encompassed under the same umbrella of executive function, the degree of task complexity changes the neural resources allocated to that task. Thus, the ability to adapt vocal motor commands may depend on task complexity. Lastly, these results may differ from Scheerer, et al. [20] due to timing differences: the two conditions in Scheerer, et al. [20] were separated by one week, whereas the conditions in the current study were completed in the same visit. Because the reliability of f_0 responses to altered auditory feedback has not been validated, it is unclear if this relatively large break between sessions impacts the study effects.

Surprisingly, from a visual inspection of the data (Figure 2), there seems to be a small difference in adaptation between conditions towards the end of the after-effect phase. The lack of an interaction effect between condition and phase could be explained by the high variability in participant data (congruent after-effect SD = 74, incongruent after-effect SD = 79). Adaptation responses are often variable and thus adds difficulty in capturing statistical changes between conditions. Group variability is a common factor experienced by many research groups working with altered auditory feedback paradigms. One purpose of this study was to determine if other neural systems are impacting the results of these paradigms, with the long-term goal of understanding if these peripheral processes contribute to the variability of responses in both typical and clinical populations. Although we did not find significant changes in adaptation between conditions, more research is necessary to interpret responses with a more wholistic view and to inform current models of speech motor control.

5. REFERENCES

- [1] S. Het, N. Rohleder, D. Schoofs, C. Kirschbaum, and O. T. Wolf, "Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'," *Psychoneuroendocrinology*, vol. 34, no. 7, pp. 1075-86, Aug 2009.
- [2] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76-81, 1993.
- [3] T. Heponiemi, N. Ravaja, M. Elovainio, P. Naatanen, and L. Keltikangas-Jarvinen, "Experiencing positive affect and negative affect during stress: relationships to cardiac reactivity and to facial expressions," *Scand J Psychol*, vol. 47, no. 5, pp. 327-37, Oct 2006.
- [4] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Jones & Bartlett Learning, 2020.
- [5] R. W. Hamill, R. E. Shapiro, and M. A. Vizzard, "Peripheral autonomic nervous system," in *Primer on the autonomic nervous system*: Elsevier, 2012, pp. 17-26.
- [6] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
- [7] H. D. Critchley, "Neural mechanisms of autonomic, affective, and cognitive integration," *Journal of comparative neurology*, vol. 493, no. 1, pp. 154-166, 2005.
- [8] H. D. Critchley *et al.*, "Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence," *Brain*, vol. 126, no. 10, pp. 2139-2152, 2003.
- [9] H. D. Critchley, J. Tang, D. Glaser, B. Butterworth, and R. J. Dolan, "Anterior cingulate activity during error and autonomic response," *Neuroimage*, vol. 27, no. 4, pp. 885-895, 2005.
- [10] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the 24th Australian computer-human interaction conference*, 2012, pp. 420-423.
- [11] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410-417, 2009.
- [12] J. L. Harbluk, Y. I. Noy, and M. Eizenman, "The impact of cognitive distraction on driver visual behaviour and vehicle control," 2002.
- [13] S. Janssen *et al.*, "Validation of the Auditory Stroop Task to increase cognitive load in walking tasks in healthy elderly and persons with Parkinson's disease," *PloS one*, vol. 14, no. 8, 2019.
- [14] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.
- [15] R. D. Seidler, Y. Kwak, B. W. Fling, and J. A. Bernard, "Neurocognitive mechanisms of error-based motor learning," in *Progress in motor control*: Springer, 2013, pp. 39-60.
- [16] B. Parrell, A. C. Lammert, G. Ciccarelli, and T. F. Quatieri, "Current models of speech motor control: A control-theoretic overview of architectures and properties," *J Acoust Soc Am*, vol. 145, no. 3, p. 1456, Mar 2019.
- [17] J. A. Tourville and F. H. Guenther, "The DIVA model: A neural theory of speech acquisition and production," *Lang Cogn Process*, vol. 26, no. 7, pp. 952-981, Jan 1 2011.
- [18] B. J. Kroger, J. Kannampuzha, and C. Neuschaefer-Rube, "Towards a neurocomputational model of speech production and perception," (in English), *Speech Communication*, vol. 51, no. 9, pp. 793-809, Sep 2009.
- [19] J. F. Houde and M. I. Jordan, "Sensorimotor adaptation in speech production," *Science*, vol. 279, no. 5354, pp. 1213-1216, 1998.
- [20] N. E. Scheerer, A. K. Tumber, and J. A. Jones, "Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback," *Journal of neurophysiology*, vol. 115, no. 2, pp. 826-832, 2016.
- [21] S. Ishihara, *Ishihara's tests for colour-deficiency*. Kanehara, 1996.
- [22] M. Berg, M. Fuchs, K. Wirkner, M. Loeffler, C. Engel, and T. Berger, "The Speaking Voice in the General Population: Normative Data and Associations to Sociodemographic and Lifestyle Factors," (in English), *Journal of Voice*, vol. 31, no. 2, Mar 2017.
- [23] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," 2017.
- [24] H. R. Weerathunge, D. Abur, N. M. Enos, K. M. Brown, and C. E. Stepp, "Auditory-motor perturbations of voice fundamental frequency: feedback delay and amplification," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 9, pp. 2846-2860, 2020.
- [25] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott. Int.*, vol. 5, no. 9, pp. 341-345, 2001.
- [26] R. A. Lester-Smith *et al.*, "The relation of articulatory and vocal auditory-motor control in typical speakers," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 11, pp. 3628-3642, 2020.
- [27] T. A. Burnett and C. R. Larson, "Early pitch-shift response is active in both steady and dynamic voice pitch control," *The Journal of the Acoustical Society of America*, vol. 112, no. 3, pp. 1058-1063, 2002.
- [28] C. R. Larson, T. A. Burnett, J. J. Bauer, S. Kiran, and T. C. Hain, "Comparison of voice F0 responses to pitch-shift onset and offset conditions," *The Journal of the Acoustical Society of America*, vol. 110, no. 6, pp. 2845-2848, 2001.
- [29] J. A. Tourville, K. J. Reilly, and F. H. Guenther, "Neural mechanisms underlying auditory feedback control of speech," *Neuroimage*, vol. 39, no. 3, pp. 1429-1443, 2008.