

# Comparing speech and nonspeech context effects across timescales in coarticulatory contexts

Navin Viswanathan<sup>1</sup> · Damian G. Kelty-Stephen<sup>2</sup>

Published online: 13 November 2017  
© The Psychonomic Society, Inc. 2017

**Abstract** Context effects are ubiquitous in speech perception and reflect the ability of human listeners to successfully perceive highly variable speech signals. In the study of how listeners compensate for coarticulatory variability, past studies have used similar effects speech and tone analogues of speech as strong support for speech-neutral, general auditory mechanisms for compensation for coarticulation. In this manuscript, we revisit compensation for coarticulation by replacing standard button-press responses with mouse-tracking responses and examining both standard geometric measures of uncertainty as well as newer information-theoretic measures that separate fast from slow mouse movements. We found that when our analyses were restricted to end-state responses, tones and speech contexts appeared to produce similar effects. However, a more detailed time-course analysis revealed systematic differences between speech and tone contexts such that listeners' responses to speech contexts, but not to tone contexts, changed across the experimental session. Analyses of the time course of effects within trials using mouse tracking indicated that speech contexts elicited fewer *x*-position flips but more area under the curve (AUC) and maximum deviation (MD), and they did so in the slower portions of mouse-tracking movements. Our results indicate critical differences between the time course of speech and nonspeech context effects and that general auditory explanations, motivated by their apparent similarity, be reexamined.

**Keywords** Speech perception · Similarity · Temporal processing

Human listeners achieve robust perception of a highly variable speech signal by demonstrating context-sensitive flexibility in speech perception. Compensation for coarticulation (CfC) is the observation that listeners' perception of a given segment depends on the nature of surrounding segments (Mann, 1980; Mann and Repp, 1981). For instance, listeners classifying a [da]–[ga] continuum report more “g” responses after segment [al] than after [aɪ]. We focus on two prominent but conflicting, explanations of this finding. From an ecological account of speech perception, listeners perceive vocal gestures that produce the acoustic signal and are attuned to acoustic consequences of coarticulation (Fowler, Brown, & Mann, 2000). Listeners report more “g” responses after [al] than [aɪ] because they perceive ambiguous steps of stop continuum as [ga] pulled forward by the frontal tongue-tip gesture of [al] but not by the posterior tongue body constriction of [aɪ]. An alternative explanation is that CfC reflects general auditory processes that respond to differences in spectral relations without explicit reference to speech production (Diehl, Lotto, & Holt, 2004). Because segments [al] and [ga] have relatively high third-formant (F3) frequencies compared to [aɪ] and [da], *spectral contrast* leads listeners to interpret the F3 of the intermediate stops as being relative lower (and more “ga”-like) after hearing a high F3 in [al] than after the low F3 in [aɪ]. Critically, CfC arises solely due to the spectral relations between the precursor and the target syllables.

Strong support for spectral contrast comes from the consistent replication of similar effects of speech and nonspeech analogues across different coarticulatory contexts (Holt, 1999, 2005, 2006; Lotto & Kluender, 1998). However, other studies pose issues for a general auditory framework. For

---

✉ Navin Viswanathan  
navin@ku.edu

<sup>1</sup> University of Kansas, Lawrence, KS, USA

<sup>2</sup> Grinnell College, Grinnell, IA, USA

instance, it is unclear whether spectral contrast conditions exist in natural speech because tone analogues fail to produce any effects when matched along critical dimensions to the speech formants they are meant to represent (Viswanathan, Fowler, & Magnuson, 2009). CfC can occur in a direction opposite to spectral contrast (Viswanathan, Magnuson, & Fowler, 2010), or in the absence of spectral contrast (Silverman, 1986; Viswanathan & Stephens, 2016).

An important limitation of prior studies demonstrating similar effects of speech and nonspeech analogues is that this apparent similarity rests on end-state target judgments using simple, time-scale-collapsing, button-press responses. Such responses both collapse within-trial decision-process variability by only considering endpoint judgments as well as across-trial variability by averaging response measures across successive trials. The ecological explanation of speech perception is one of attunement such that perceivers explore an informational space over the timescale of an entire experiment and settle into facility with specifying properties of an event (Withagen & Michaels, 2005). Within trials, the perception of target segments with identical acoustic manifestation may be affected by the coarticulatory consequences of preceding segments (e.g., Fowler & Smith, 1986). Consequently, identifying target segments involve both fast perceptual processes attending to the segmental acoustics, slower ones attuning to effects of preceding contexts, and the slowest ones that reflect attunement across the experimental session. In comparison, general auditory processes for successful CfC are based on contrast effects of the preceding context within the trial. Indeed, other studies using longer speech such as ones matched to sentence duration (e.g., Stilp, Anderson, & Winn, 2015), tone sequences (e.g., Holt, 2005), and other nonspeech analogues (e.g., Sjerps, Mitterer, & McQueen, 2011) demonstrate longer-term contrast effects. Because our focus is on evaluating CfC explanations, we will compare the time course of the effects of single, syllable-length tone precursors and speech precursors on the categorization of following target speech. Consequently, our results do not speak directly to the longer-term contrast effects demonstrated by the aforementioned studies.

In this article, we evaluate whether time-course measures of target identification reveal systematic differences between perceptual effects of speech and tone contexts. Within trials we will use mouse tracking to study speech and nonspeech contexts shown to produce similar effects in end-state judgments of the target continuum. We also will track how these effects change across the experiment. By comparing these context effects against a no-context baseline, we will test whether the similarity of speech and nonspeech effects holds across perceptual timescales ranging from within single trials to across the experimental session. A two-pronged approach to quantifying mouse-tracking behavior will allow testing parallel hypotheses addressing relationships among context type

(context), precursor formant structure (precursor), and processing speed. If we take the mouse-tracking behavior to be a portrayal of online perceptual processing, the balancing of information through perceptual processing should manifest in the time course of mouse-tracked movements. Novel analyses allows quantifying the informational content of mouse-tracking behavior and to do so separately for the fast, ballistic movements following participants committing to a direction/decision and for the slow, pause-laden hoverings as participants make their choices (Calcagni, Lombardi, & Sulpizio, 2017). Following an old tradition in cognitive science (e.g., Miller, 1956) that remains current in cognitive neuroscience (Fan et al., 2014), we can quantify informational content in Shannon's (1948) information-theoretic terms. Calcagni et al.'s method converts  $x$ - $y$  coordinates of mouse-tracking data into polar coordinates, computes histograms for angular directions during "micropause" fixation-like movements and for angular directions during fast, ballistic saccade-like movements, and computes the Shannon entropy for each histogram. Shannon entropy is a measure of variability that expresses how uniformly a random process spreads itself over a wide range of values. Shannon entropy of recorded movements will increase as recorded movements exemplify a both a greater variety of angle values and greater similarity in the frequency of each value. Conversely, Shannon entropy increases as recorded movements exhibit mostly a limited range of angles and much fewer of other alternative angles.

Our first hypothesis is about the interaction of speech context with precursor and how this changes across the experiment. Specifically, we expect that effects of context on RT as well as on the judgment of the target stimulus will change across the experiment preferentially for speech contexts. Besides testing expected increases in RT for intermediary, ambiguous tokens from the continuum, we expect differential CfC effects (i.e., significant Context  $\times$  Precursor interaction) controlling for slow-fast differences in information processing, CfC will express itself with greater strength over the course of the experiment, with both RT and judgments showing progressively greater sensitivity to precursor formant structure in the speech context (Hypothesis 1).

Second, we take a new look at "geometric" responses ( $x$ -position flips, area under the curve [AUC], and maximum deviations [MD] of mouse trajectories from shortest path) to describe within-trial progression of identification responses (Freeman & Dale, 2013). All of these geometric measures capture aspects of uncertainty in the mouse-tracking movements and should increase for the ambiguous, intermediary-step target stimuli.  $X$ -position flips reflect vacillation between the left or the right response choices. MD is the maximal distance away from shortest (linear) path, and AUC is the area under the path's curve. Convexity or angularity of the mouse path defines how these measures might increase together at similar or different rates. Despite portraying what should look

like the embodied decision process of a participant struggling with an ambiguous stimulus, the geometric measures are insensitive to both the heterogeneity of speeds and the information content in mouse-tracking movements. Besides expecting that all three geometric measures should increase for the ambiguous, intermediary-step target stimuli, we expect the slow component of mouse-tracking movements to predict uncertainty in the response, specifically for speech contexts and the CfC-focused interaction Context  $\times$  Precursor, to predict gradually less uncertainty indexed by each geometric measure (Hypothesis 2). That is, because the context's effects on target identification will depend on slower processes that span longer timescales, systematic differences between speech and nonspeech contexts exist within trials should manifest as differences mainly in slower movements specifically reflecting perceptual processing of slower scale coarticulatory information in speech contexts.

## Method

### Participants

Thirty-six native English-speaking University of Kansas undergraduates who reported normal hearing and corrected or normal vision participated in the study after providing informed consent. Experimenters randomly assigned each to one of three groups: no-context, speech-context, or tone-context conditions.

### Materials

An 11-step continuum of resynthesized CV syllables varying in F3-onset frequency in 100-Hz steps, from 1800 to 2800 Hz, changing linearly to 2500-Hz steady-state offsets, varying perceptually from [da] to [ga] was used (see Viswanathan et al., 2009). The entire continuum had identical first, second, and fourth formants (i.e., shifting from 500 Hz to 800 Hz, shifting from 1600 Hz to 1200 Hz, and remaining at 3500 Hz, respectively). Each CV syllable was 215-ms long. Speech contexts were natural tokens of [al] and [aɪ], with matching 375-ms duration and intensity. Critical F3 offsets were approximately 2600 Hz for [al] and 1820 Hz for [aɪ]. Tone contexts were duration-matched and intensity-matched tone analogues of [al] and [aɪ] mimicking respective F3-offset frequencies (cf. Viswanathan et al., 2009). Throughout, 50 ms of silence separated precursor and target.

### Procedure

In a two-alternative forced-choice task, participants clicked the computer-mouse cursor over a “Start” box at the screen's lower center to play stimuli and indicated judgments by

clicking either the top left or top right of the screen labeled “ga” or “da” (location counterbalanced across participants).

After 10 practice trials presenting only [da] and [ga] end points in random order, participants completed 176 trials (16 repetitions of the 11-step stop continuum), judging members of stop continuum in the no-context group. The speech-context and tone-context group judged these stops in disyllable sequences and in tone-speech sequences, respectively. The onset of the next trial was controlled by participants who had an option to take breaks between trials. The overall experiment lasted under 30 minutes. No feedback was provided.

### Analysis

**Mouse-tracking analysis** E-Prime® software collected ( $x, y$ ) mouse-tracking data at approximately 58 Hz. The R package “mousetrap” calculated  $x$ -position flips, AUC, and MD statistics from mouse-tracking trajectories (Kieslich, Wulff, Henninger, & Haslbeck, 2017). Calcagni et al.'s (2017) informational-entropy measures  $\psi$  and  $\xi$  quantify entropy for the entire trajectory and only for fast movements, respectively. Consequently, we refer to  $\psi$  as capturing slow entropy because the  $\xi$  encoding fast-movement entropy was always included in the model.

**Modeling of RT, “ga” selection, and “mousetrap” measures** Linear mixed-effect modeling was used to test both hypotheses because RT and mousetrap measures varied continuously. Poisson mixed-effect modeling was used to test tested Hypothesis 1 because individual “ga” selections were dichotomous, and cumulative “ga” selections across 176 trials were better modeled as a count variable than as a continuous variable. Whereas logistic modeling uses logit links to model odds of a single dichotomous events, Poisson modeling uses log links to model the marginal probability of one more event. Poisson modeling allowed trial effects to explicitly model actual sequence of the dependent variable and allowed random-effects structure without compromising convergence in logistic models.

Throughout, mixed-effect modeling used random-effect intercepts per participant and random slopes on linear and quadratic effects of Step. Random-effect and fixed-effect instances of linear and quadratic effects used orthogonal polynomials to eliminate correlation between them.<sup>1</sup> In addition to predictor Step, other predictors were Context (i.e., Speech or Tone), Precursor (Low or [aɪ] = 1, High or [al] = 2), total entropy  $\psi$ , and fast entropy  $\xi$ . All terms not including explicit interaction with Context refer to effects in the no-context case,

<sup>1</sup> Alternative modeling showed random effect for precursor-rendered interactions Context  $\times$  Step (Linear) and Precursor  $\times$  Step (Linear) nonsignificant only for the Poisson model of “ga” selections. No other model changed significantly with precursor random effect.

that is, beyond the “simple replication” portion of Results, all modeling addresses all data including no-context cases. Context (None) is the control case of the categorical variable Context, and it is customary for modeling to omit explicit listing of the control case (e.g., Bates, Maechler, Bolker, & Walker, 2015). Just as full-factorial structure allows ANOVA models to estimate that three-way interaction effects are not artifactually attributable to main effects or to two-way interactions, regression modeling to test present work’s hypotheses of higher-order interactions must include all lower-order terms composing the interaction (e.g., the main effects and lower-order interactions). These components are necessary, but not all of them inform experimenters’ hypotheses, and not all will be sensibly interpretable except as adequate control for terms that are pertinent to the hypotheses of interest. For instance, our interest in the quadratic effect of Step comes from the expectation that middle items on the Step continuum, but we include the linear effect of Step as a control term because the quadratic term (i.e.,  $\text{Step}^2$ ) is the interaction of the linear effect Step with itself (i.e.,  $\text{Step}^2 = \text{Step} \times \text{Step}$ ), and so any true quadratic effect requires also modeling the main-effect linear term. In any traditional experimental design, we recruit a control condition as well as a treatment condition because experimenters are interested in the difference. Notably, experimenters are rarely concerned with the scores or measures from the control group, except as a comparison group. So in the same way, we focus on the hypothesized quadratic as the evidence of a difference from the comparison-group-like linear term, and we generally omit attention to the control effect of Step.<sup>2</sup>

## Results

### Simple replication

Similar to past CfC studies (Viswanathan, Magnuson, & Fowler, 2014), we submitted logit-transformed response proportions from the no-context condition to a one-factor (Step) 11-level ANOVA and submitted a logit-transformed response proportion from the speech and tone contexts to a 2 (Context)  $\times$  11 (Step) ANOVA. Effects of step were consistent across all conditions ( $p < .001$ ;  $\eta_p^2 > 0.67$ ). We replicated precursor effects in both speech and tone contexts (average shifts = 10.42% and 12.03%;  $p = .02$  and  $.04$ ,  $\eta_p^2 = .59$  and  $.54$ , respectively; see Fig. 1).

<sup>2</sup> We offer more extended remarks as an online appendix, reviewing select aspects of regression modeling in general but also a table of explanation of what each effect means, as well as indication of each effect’s status as either a control or substantive effect ([https://sites.google.com/site/foovian/VKS\\_Appendix.pdf](https://sites.google.com/site/foovian/VKS_Appendix.pdf)).

### Across trial effects: response time and identification became more precursor-dependent, with speech but not tone contexts

Models of RT and Identification show a curious similarity in that both show that response time and response change with Trial. Importantly, in speech, but not in tone, contexts, the effect of the precursor changes over trials.

#### Response time

After logarithmically transforming RT to reduce skew, we fit the linear mixed-effect model (see Table 1). RT increased marginally over middle steps (i.e., for Step [Quadratic]) indicating increased uncertainty for the ambiguous steps and showed more rapid responses over later trials. However, speech contexts led RT to decrease more slowly over trials (Context (Speech)  $\times$  Trial). This shallower decrease depended on precursor with [a] slowing participants’ response less than [a]. Both contexts produced linear decrease of RT with greater Step, particularly for [a] or its tone analogue. Note that the dependence of RT on trial as well as on precursor depends strongly on the speech context. Temporal effects of tone context was not significantly different from the no-context baseline.

#### Identification responses

To incorporate the effect of trials, we used Poisson-modeled effects on the logarithmic probability of a participant adding one more “ga” selection. Here, too, we found that with longer experience with speech contexts, the effect of precursor grew stronger. Specifically, the interaction of Context (Speech) with Trial specified a decrease over trials, and for the [a] precursor, the positive effect of Context (Speech)  $\times$  Precursor  $\times$  Trial leaves this change with Trial negative. However, for [a] (Precursor = 2), this interaction becomes positive (i.e.,  $-.01 + 2 (.0059) = .0018$ ). Hence, longer experience with speech context gives the precursor greater effect on the consequent judgment.

Models of RT and cumulative “ga” selections show independent effects of speech versus tone context difference and of CfC (see Fig. 2). Specifically, the model predictions show that, across cumulative trials in the experiment, the differences due to context dwindle and that differences due to context-by-precursor interactions become more prominent.

### Within-trial effects: change in mouse-tracking measures with speech contexts but not tone contexts at both fast and slower timescale perceptual responses

Subsequent models of within-trial progression of phonetic judgment (see Table 2) show ambiguous stimuli amplifying geometrical mouse-tracking measures classically associated with uncertainty. A striking feature of the model of  $x$ -position

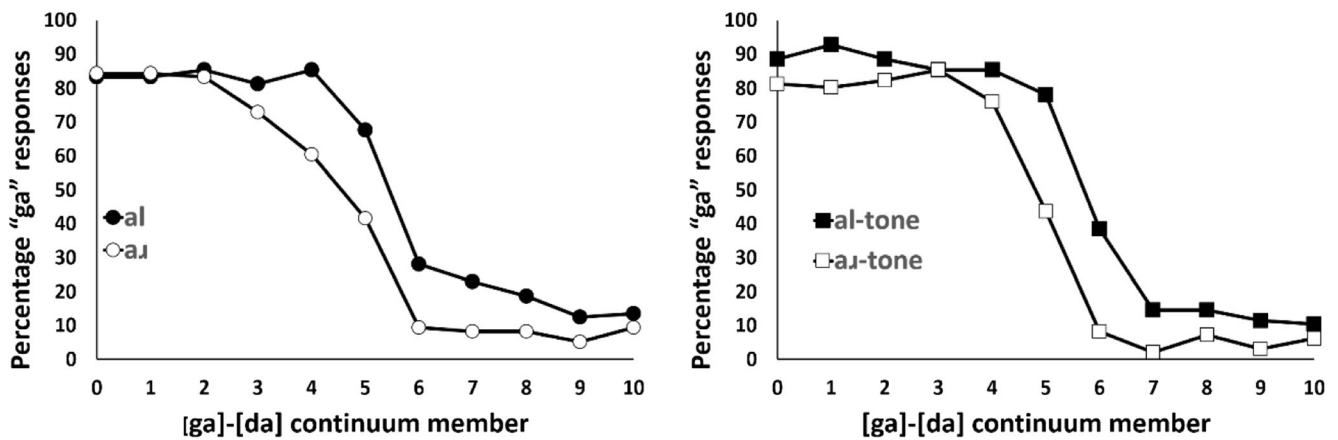


Fig. 1 Analysis of end-state judgments demonstrate that speech (left) and tone contexts (right) appear to produce similar shifts on identification

flips is its similarity with both the RT and ga-selection models. Namely, all three of these models show a change with Trial, Precursor, or Precursor × Trial for participants experiencing speech but not tone contexts (see Fig. 3). Throughout, Context (Speech) × Precursor and Context (Speech) × Trial having the same sign, opposite that of Trial, Context (Speech) × Trial, and Context (Speech) × Precursor × Trial.

**Trial-dependent reduction in x-position flips with speech context but not tone context**

X-position flips increased for the intermediate and most ambiguous values of Step (Step (Quadratic)). Interactions of  $\psi$  with Trial × Context (Speech), all lower-order interactions,

**Table 1** Across Trial effects: Linear regression model for logarithmically transformed response time (logRT). Terms in bold indicate critical predictors that were significant

Predictor	B	SE	p
Intercept	6.53	.08	<.0001
Context (Speech)	-.22	.13	.09
Context (Tone)	-.21	.13	.11
Step (Linear)	-1.58	1.15	.17
<b>Step (Quadratic)</b>	<b>-1.60</b>	<b>.87</b>	<b>.06</b>
Precursor	-.02	.05	.65
<b>Trial</b>	<b>-.0040</b>	<b>.0002</b>	<b>&lt;.0001</b>
Context (Speech) × Step (Linear)	4.56	2.49	.07
Context (Speech) × Step (Quadratic)	-1.38	2.25	.54
Context (Tone) × Step (Linear)	6.40	2.51	<.05
Context (Tone) × Step (Quadratic)	-5.16	2.27	<.05
<b>Precursor × Step (Linear)</b>	<b>-3.45</b>	<b>1.25</b>	<b>&lt;.01</b>
Precursor × Step (Quadratic)	1.74	1.25	.16
Precursor × Trial	.0002	.0005	.73
<b>Context (Speech) × Trial</b>	<b>.0034</b>	<b>.0007</b>	<b>&lt;.0001</b>
Context (Tone) × Trial	.0009	.0008	.26
Context (Speech) × Precursor	.11	.07	.11
<b>Context (Speech) × Precursor × Trial</b>	<b>-.0014</b>	<b>.0007</b>	<b>&lt;.05</b>

and  $\psi$  main effects contributed significantly to model fit for x-position reversals,  $\chi^2(12) = 878.01, p < .0001$  (see Table 3). Briefly, x-position flips decrease with greater Trial and with greater slow entropy for participants experiencing a speech context. So, the participants’ trajectory in the decision process shows significantly fewer indecisive reversals in terms of the lateral direction of the mouse movements.

**Change of AUC and MD in the fast and slower timescale perceptual responses with speech but not tone**

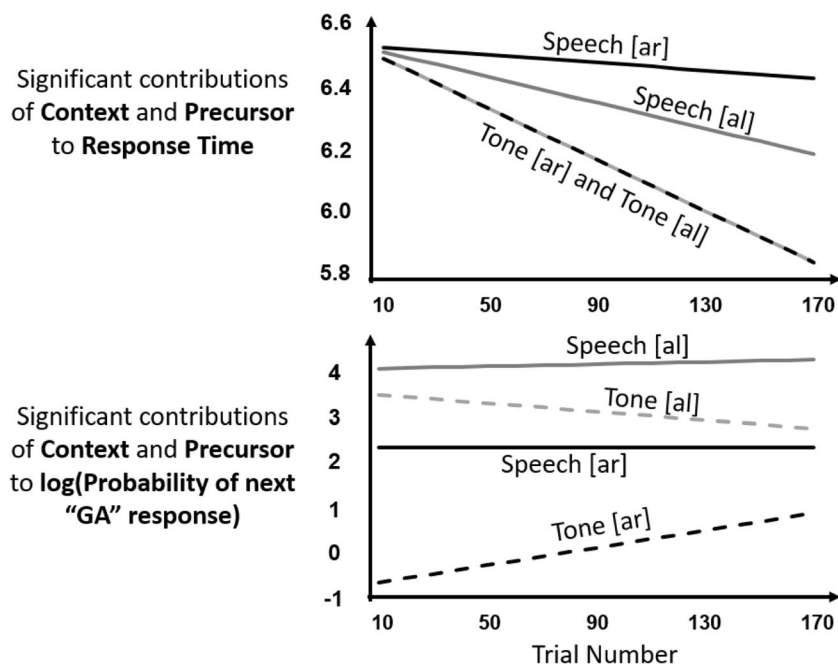
The AUC model (see Table 4) found a linear effect of Step, main effects  $\psi$  and  $\xi$ , and interaction  $\psi \times \text{Trial} \times \text{Context (Speech)}$ . Hence, AUC decreases with stimuli sounding more like “da” than “ga.” Mouse trajectories with greater  $\psi$  show larger AUC, and this  $\psi$  dependence of AUC only increases with trial for participants experiencing speech contexts.

The MD model (see Table 5) returned significant effects for  $\psi$ ,  $\psi \times \text{Trial}$ , and  $\psi \times \text{Trial} \times \text{Context (Speech)}$ , as well as for  $\xi$  and  $\xi \times \text{Trial} \times \text{Context (Speech)}$ . MD increased with Trial. As in the AUC model, participants experiencing speech contexts increased MD with successive Trials. Participants experiencing either no context or tone context showed a modest growth of MD with successive Trials, tempered by  $\psi$ .

Note that the effects that diminished x-position flips are some of the same that increase MD and AUC. Figure 2 schematizes these concurrent effects on stereotypical mouse-tracking trajectories. These common effects indicate that, no matter the geometric mouse-tracking measure we consult, the change in effects of speech context are moderated by slow perceptual processing in all cases.

**Discussion**

We hypothesized that CfC would show a change in the use of speech relative to tone and no contexts over successive trials



**Fig. 2** Model predictions using significant coefficients from regression models of RT (top panel) and logarithmic probability of cumulative “ga” selection (bottom panel) across experiment trials. RT decreases similarly for all tone-context stimuli but more slowly for speech-context stimuli, with quicker decrease for [al] than [ai] precursors. Probability of “ga”

selections shows diminishing difference of speech-context stimuli from tone-context stimuli, diminishing precursor differences within tone-context stimuli, but increasing precursor differences within speech-context stimuli

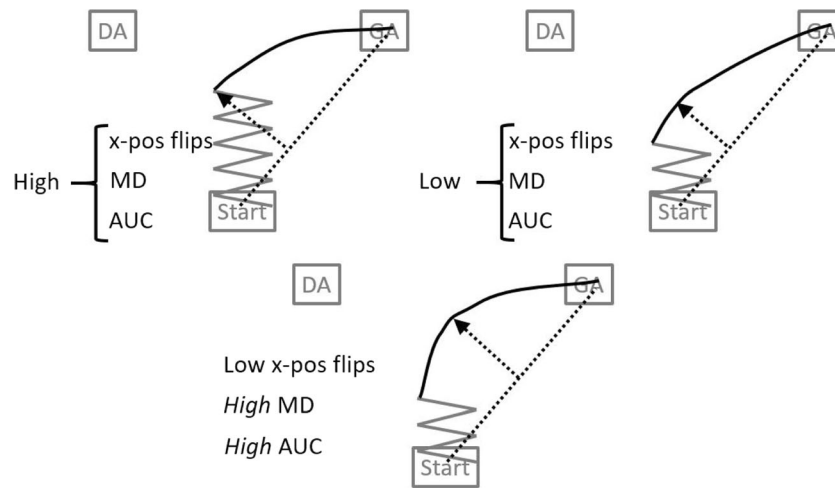
(Hypothesis 1) and that CfC would exhibit sensitivity to the speech or tone context, especially in the slower components of mouse-tracking trajectories (Hypothesis 2). Results supported both hypotheses.

**Table 2** Poisson regression model for cumulative identifications of “ga” beyond the no-context baseline condition. Terms in bold indicate critical predictors that were significant

Predictor	<i>B</i>	<i>SE</i>	<i>p</i>
Intercept	2.39	.11	<.0001
Context (Speech)	2.33	.15	<.0001
Context (Tone)	−.79	.15	<.0001
Step (Linear)	−.09	.36	.80
Step (Quadratic)	.0035	.32	.91
Precursor	1.15	.0016	<.0001
Trial	.01	$7.73 \times 10^{-7}$	<.0001
Context (Speech) × Step (Linear)	−2.21	.66	<.001
Context (Speech) × Step (Quadratic)	.42	.62	.50
Context (Tone) × Step (Linear)	−1.47	.74	<.05
Context (Tone) × Step (Quadratic)	.58	.71	.42
Precursor × Step (Linear)	1.30	.33	<.0001
Precursor × Step (Quadratic)	−.57	.33	.08
Precursor × Trial	−.0047	.0001	<.0001
<b>Context (Speech) × Trial</b>	<b>−.01</b>	<b>.0002</b>	<b>&lt;.0001</b>
Context (Tone) × Trial	−.0003	.0003	.19
Context (Speech) × Precursor	−1.80	.02	<.0001
<b>Context (Speech) × Precursor × Trial</b>	<b>.0059</b>	<b>.0002</b>	<b>&lt;.0001</b>

We replicated similar effects of speech and nonspeech tones on end-state categorization of subsequent target segments in mouse-click responses. Across trials, we found that the effects of speech contexts on the target played an ever stronger role in “ga” or “da” decisions. This growing role of speech context required more consideration before making the decision, and so whereas RT in the no-context condition tapered off with successive trials, this growing role of speech manifested in a slower decrease in RTs for the speech-context condition, while RTs for the tone-context condition decreased with trial no differently from the no-context case. The slower decrease in response time for the speech-context condition across trials depended on the precursor, with [al] precursors showing relatively more similar RT profile to no-context condition. Similar to across trial effects, geometric mouse-tracking measures of judgment uncertainty showed that the effect of the speech but not the tone contexts changed over time. Hence, listeners gradually attune to the effects of the speech but not tone contexts with continued exposure.

The current study again suggests apparent similarity when only considering end-state judgments. However, differences between the speech and tone contexts are revealed when more sensitive time-course methods are used. Specifically, we found that repeated experience with the CfC task significantly changed how participants made use of the speech context both in terms of response time and in terms of decision. Furthermore, all models showed the same set of effects indicating that, throughout mouse-click responses, successive



**Fig. 3** Schematics illustrating how the geometric mouse-tracking measures of uncertainty (*x*-position flips, MD, and AUC) might change in the same directions (top panels), whether with all measures becoming

high (top left) or low (top right). However, present results indicated that the same effects decreased *x*-position flips but increased MD and AUC. The bottom panel schematizes how this pattern of effects coexist

trials in this task reveals new sensitivity to speech contexts that go beyond responses to tones capturing F3 offsets. As indicated earlier, the current study focuses on whether CfC effects

produced by speech are due to the spectral contrast produced by the immediately preceding segment. The current study, in combination with previous reviewed studies (Viswanathan

**Table 3** Within-trial effects: Linear regression model for *x*-position flips incorporating entropy parameters. Terms in bold indicate critical predictors that were significant

Predictor	<i>B</i>	<i>SE</i>	<i>p</i>
Intercept	.30	.08	<.01
Context (Speech)	-.48	.15	<.01
Context (Tone)	.0057	.15	.97
Step (Linear)	-4.02	2.10	.06
<b>Step (Quadratic)</b>	<b>-2.82</b>	<b>1.24</b>	<b>&lt;.05</b>
Precursor	-1.26	.07	.06
Trial	-.0011	.0005	<.05
<b>ψ</b>	<b>.89</b>	<b>.06</b>	<b>&lt;.0001</b>
<b>ξ</b>	<b>-.51</b>	<b>.04</b>	<b>&lt;.0001</b>
Context (Speech) × Step (Linear)	3.55	3.94	.37
Context (Speech) × Step (Quadratic)	4.65	3.13	.14
Context (Tone) × Step (Linear)	2.25	3.96	.57
Context (Tone) × Step (Quadratic)	1.43	3.16	.65
Precursor × Step (Linear)	.18	1.72	.91
Precursor × Step (Quadratic)	-.10	1.72	.95
Precursor × Trial	.0012	.0007	.07
<b>Context (Speech) × Trial</b>	<b>.0040</b>	<b>.0012</b>	<b>&lt;.001</b>
Context (Tone) × Trial	-.0012	.0012	.35
<b>Context (Speech) × Precursor</b>	<b>.24</b>	<b>.09</b>	<b>&lt;.01</b>
ψ × Trial	.0001	.0005	.87
ψ × Context (Speech)	.11	.08	.17
ψ × Context (Tone)	.11	.08	.17
<b>Context (Speech) × Precursor × Trial</b>	<b>-.0024</b>	<b>.0009</b>	<b>&lt;.01</b>
<b>ψ × Trial × Context (Speech)</b>	<b>-.0017</b>	<b>.0008</b>	<b>&lt;.05</b>
ψ × Trial × Context (Tone)	-.0008	.0008	.34

**Table 4** Linear regression model for AUC incorporating entropy parameters. Terms in bold indicate critical predictors that were significant

Predictor	<i>B</i>	<i>SE</i>	<i>p</i>
Intercept	-2.40 × 10 <sup>3</sup>	1.14 × 10 <sup>3</sup>	<.05
Context (Speech)	-1.55 × 10 <sup>3</sup>	2.42 × 10 <sup>3</sup>	.52
Context (Tone)	-1.51 × 10 <sup>3</sup>	2.43 × 10 <sup>3</sup>	.53
Step (Linear)	-6.07 × 10 <sup>4</sup>	2.92 × 10 <sup>4</sup>	<.05
Step (Quadratic)	-3.53 × 10 <sup>4</sup>	2.33 × 10 <sup>4</sup>	.13
Precursor	1.77 × 10 <sup>3</sup>	1.25 × 10 <sup>3</sup>	.16
Trial	15.77	10.09	.12
<b>ψ</b>	<b>7.42 × 10<sup>3</sup></b>	<b>1.15 × 10<sup>3</sup></b>	<b>&lt;.0001</b>
<b>ξ</b>	<b>-3.55 × 10<sup>3</sup></b>	<b>7.62 × 10<sup>2</sup></b>	<b>&lt;.0001</b>
Context (Speech) × Step (Linear)	8.87 × 10 <sup>4</sup>	6.42 × 10 <sup>4</sup>	.17
Context (Speech) × Step (Quadratic)	4.54 × 10 <sup>2</sup>	5.91 × 10 <sup>4</sup>	.99
Context (Tone) × Step (Linear)	6.43 × 10 <sup>4</sup>	6.46 × 10 <sup>4</sup>	.32
Context (Tone) × Step (Quadratic)	-1.14 × 10 <sup>4</sup>	5.97 × 10 <sup>4</sup>	.85
Precursor × Step (Linear)	-7.7 × 10 <sup>3</sup>	3.26 × 10 <sup>4</sup>	.81
Precursor × Step (Quadratic)	2.55 × 10 <sup>4</sup>	3.26 × 10 <sup>4</sup>	.43
Precursor × Trial	-6.11	12.57	.63
Context (Speech) × Trial	1.68	22.78	.94
Context (Tone) × Trial	.78	23.51	.97
Context (Speech) × Precursor	-6.57 × 10 <sup>2</sup>	1.72 × 10 <sup>3</sup>	.70
ψ × Trial	-13.25	10.42	.20
ψ × Context (Speech)	-5.90 × 10 <sup>2</sup>	1.45 × 10 <sup>3</sup>	.68
ψ × Context (Tone)	-2.54 × 10 <sup>3</sup>	1.52 × 10 <sup>3</sup>	.10
Context (Speech) × Precursor × Trial	.40	17.18	.98
<b>ψ × Trial × Context (Speech)</b>	<b>30.73</b>	<b>14.44</b>	<b>&lt;.05</b>
<b>ψ × Trial × Context (Tone)</b>	<b>29.56</b>	<b>15.53</b>	<b>.06</b>

**Table 5** Linear regression model for MD incorporating entropy parameters. Terms in bold indicate critical predictors that were significant

Predictor	<i>B</i>	<i>SE</i>	<i>p</i>
Intercept	−20.13	6.50	<.01
Context (Speech)	−14.02	13.62	.30
Context (Tone)	5.50	13.65	.69
Step (Linear)	−154.13	136.99	.26
Step (Quadratic)	−219.90	128.08	.09
Precursor	3.21	6.94	.64
Trial	.15	.06	<.01
$\psi$	<b>55.31</b>	<b>6.41</b>	<b>&lt;.0001</b>
$\xi$	<b>−33.82</b>	<b>4.25</b>	<b>&lt;.0001</b>
Context (Speech) × Step (Linear)	58.57	334.93	.86
Context (Speech) × Step (Quadratic)	225.20	327.80	.49
Context (Tone) × Step (Linear)	182.69	337.47	.59
Context (Tone) × Step (Quadratic)	262.38	331.13	.43
Precursor × Step (Linear)	−38.07	181.45	.83
Precursor × Step (Quadratic)	−54.52	181.68	.76
Precursor × Trial	.02	.07	.78
Context (Speech) × Trial	.04	.13	.74
Context (Tone) × Trial	−.12	.13	.36
Context (Speech) × Precursor	6.49	9.59	.50
$\psi$ × Trial	<b>−.12</b>	<b>.06</b>	<b>&lt;.05</b>
$\psi$ × Context (Speech)	7.94	8.08	.33
$\psi$ × Context (Tone)	−9.34	8.47	.27
Context (Speech) × Precursor × Trial	−.09	.10	.33
$\psi$ × Trial × Context (Speech)	<b>.16</b>	<b>.08</b>	<b>&lt;.05</b>
$\psi$ × Trial × Context (Tone)	<b>.19</b>	<b>.09</b>	<b>&lt;.05</b>

et al., 2009; Viswanathan, Magnuson, & Fowler, 2010, 2013; Viswanathan et al., 2014), call the spectral contrast effect of CfC into question. However, these studies do not address other longer-term spectral contrast effects posited to explain listeners' accommodation to other forms of variability (e.g., vowel normalization).

In elaborating CfC responses beyond traditional button-press into continuous mouse-tracking measures, we hoped to exploit both standard geometric measures (e.g., AUC, MD) and the more-recent information-theoretic decompositions of mouse-movement entropy. Fortunately, whereas CfC has invited diverging interpretations, different mouse-tracking modeling strategies do not pose a new set of contrary perspectives. Rather, these approaches complement each other well, allowing information-theoretic measures to distinguish the contingency of speech-context effects on slower attentional processes from the robust effect of CfC in both slow and fast processing. The complementary approaches allow us to make specific conclusions about how CfC changes across and within trials (i.e., how continued experience over the entire experiment changes the slow and fast aspects of response within a single trial).

Present investigation of CfC brings into relief the richly multiscale structure of speech perception. Because the coarticulatory role of contextual segments on target-segment acoustics unfolds more slowly than the target itself, listeners must attune to information unfolding at multiple timescales in speech. Present work shows that their attunement to context is relatively trial invariant for tone contexts but appears for speech contexts over the slowest components of single-trial response with greater experience in the experimental task.

**Author note** This research was supported by NIDCD Grant R15 DC011875-01 to N.V. and N.S.F., Grant BCS-1431105 to N.V. Both authors contributed equally to all aspects of the reported research.

## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Calcagni, A., Lombardi, L., & Sulpizio, S. (2017). Analyzing spatial data from mouse tracker methodology: An entropic approach. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-016-0839-5>
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- Fan, J., Van Dam, N. T., Gu, X., Liu, X., Wang, H., Tang, C. Y., & Hof, P. R. (2014). Quantitative characterization of functional anatomical contributions to cognitive control under uncertainty. *Journal of Cognitive Neuroscience*, 26, 1490–1506.
- Fowler, C. A., Brown, J., & Mann, V. (2000). Contrast effects do not underlie effects of preceding liquid consonants on stop identification in humans. *Journal of Experimental Psychology: Human Perception & Performance*, 26, 877–888.
- Fowler, C., & Smith, M. R. (1986). Speech perception as vector analysis: An approach to the problems of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–139). Hillsdale, NJ: Erlbaum.
- Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavioral Research Methods*, 45, 83–97.
- Holt, L. L. (1999). *Auditory constraints on speech perception: An examination of spectral contrast* (Unpublished doctoral dissertation), University of Wisconsin–Madison.
- Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science*, 16, 305–312.
- Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of Acoustical Society of America*, 120, 13–2817.
- Kieslich, P. J., Wulff, D. U., Henninger, F., & Haslbeck, J. M. B. (2017). Mouseltrap: An R package for processing and analyzing mouse-tracking data (Version 3.0.0) [Computer software]. <https://doi.org/10.5281/zenodo.290283>
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602–619.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548–558.



- Miller, G. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Silverman, K. (1986). F<sub>0</sub> cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, 43, 76–92.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention Perception, & Psychophysics*, 73(4), 1195–1215.
- Stilp, C. E., Anderson, P. W., & Winn, M. B. (2015). Predicting contrast effects following reliable spectral properties in speech perception. *The Journal of the Acoustical Society of America*, 137(6), 3466–3476.
- Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review*, 16, 74–79.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception & Performance*, 36, 1005–1015.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2013). Similar response patterns do not imply identical origins: An energetic masking account of nonspeech effects in compensation for coarticulation. *Journal of Experimental Psychology: Human Perception and Performance*, 39(4), 1181–1192.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2014). Information for coarticulation: Static signal properties or formant dynamics? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1228–1236.
- Viswanathan, N., & Stephens, J. D. (2016). Compensation for visually specified coarticulation in liquid–stop contexts. *Attention, Perception, & Psychophysics*, 78(8), 2341–2347.
- Withagen, R., & Michaels, C. F. (2005). The role of feedback information for calibration and attunement in perceiving length by dynamic touch. *Journal of Experimental Psychology: Human Perception & Performance*, 31, 1379–1390.