JASA ARTICLE

The effects of variability on context effects and psychometric function slopes in speaking rate normalization[a)]

Caleb J. King,[1] Chloe M. Sharpe,[2] Anya E. Shorey,[1] and Christian E. Stilp[1,b)]

[1]*Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292, USA*

[2]*School of Psychology, Xavier University, Cincinnati, Ohio 45207, USA*

ABSTRACT:

Acoustic context influences speech perception, but contextual variability restricts this influence. Assgari and Stilp [J. Acoust. Soc. Am. **138**, 3023–3032 (2015)] demonstrated that when categorizing vowels, variability in who spoke the preceding context sentence on each trial but not the sentence contents diminished the resulting spectral contrast effects (perceptual shifts in categorization stemming from spectral differences between sounds). Yet, how such contextual variability affects temporal contrast effects (TCEs) (also known as speaking rate normalization; categorization shifts stemming from temporal differences) is unknown. Here, stimuli were the same context sentences and conditions (one talker saying one sentence, one talker saying 200 sentences, 200 talkers saying 200 sentences) used in Assgari and Stilp [J. Acoust. Soc. Am. **138**, 3023–3032 (2015)], but set to fast or slow speaking rates to encourage perception of target words as "tier" or "deer," respectively. In Experiment 1, sentence variability and talker variability each diminished TCE magnitudes; talker variability also produced shallower psychometric function slopes. In Experiment 2, when speaking rates were matched across the 200-sentences conditions, neither TCE magnitudes nor slopes differed across conditions. In Experiment 3, matching slow and fast rates across all conditions failed to produce equal TCEs and slopes everywhere. Results suggest a complex interplay between acoustic, talker, and sentence variability in shaping TCEs in speech perception.
https://doi.org/10.1121/10.0025292

(Received 22 September 2023; revised 23 February 2024; accepted 27 February 2024; published online 14 March 2024)

[Editor: Sven Mattys] Pages: 2099–2113

I. INTRODUCTION

In auditory perception, the surrounding acoustic context influences what sounds are perceived. Acoustic differences between surrounding acoustic context and target sounds are perceptually magnified, resulting in contrast effects where larger changes are perceived than are physically present. These contrast effects exist across multiple domains. First, spectral contrast effects (SCEs) arise due to differences in the spectral contents of two sounds. In a classic demonstration, Ladefoged and Broadbent (1957) found that shifting the spectrum of a context sentence influenced the perception of the following target word. When higher first formant (F_1) frequencies were emphasized in the context sentence, participants perceived the target sound to be lower-F_1 "bit." Likewise, when lower F_1 frequencies were emphasized, participants judged the same target sound to be higher-F_1 "bet." These SCEs can be calculated through the change in a particular response following different spectral contexts (e.g., percent changes in "bet" responses following the low-F_1 versus high-F_1 context sentence; n. b., other calculations are possible, such as the shift in category boundaries/50% points on psychometric functions for responses to targets following different contexts). Similar examples have been reported for perception of F_2 (Mitterer, 2006), F_3 (Holt, 2005), and spectral shape (Watkins, 1991).

A second type of auditory contrast effect results from differences in temporal properties between two sounds. These are known as temporal contrast effects (TCEs) (or speaking rate normalization). When a context sentence is spoken at a fast rate, the subsequent target is perceived as having slower or longer temporal characteristics, such as a longer duration (Reinisch and Sjerps, 2013), longer voice onset time (VOT) (Summerfield, 1981), longer formant transitions (Minifie *et al.*, 1977), or longer rise time (Repp *et al.*, 1978); conversely, a context sentence spoken at a slow rate results in the perception of shorter-duration stimulus properties. As above, these TCEs can be calculated through the change in a particular response following different temporal contexts (e.g., percent changes in "deer" responses following the slow versus fast context sentence). Listeners form expectations based on speaking rate, which in turn affects the perception of words and the boundaries between them (Dilley and Pitt, 2010). Both spectral and temporal contrast effects affect how listeners perceive speech (see Stilp, 2020a, for review).

While both SCEs and TCEs produce a contrastive outcome by shifting perception away from acoustic properties of

[a)]Portions of these results were presented at the 182nd and 184th Meeting of the Acoustical Society of America.

[b)]Current address: 317 Life Sciences Building, Louisville, KY 40292, USA. Email: christian.stilp@louisville.edu surrounding sounds, these contrasts appear to be subserved by distinct neural mechanisms. SCEs are proposed to be produced by neural adaptation, primarily (but not exclusively) in the auditory periphery (Stilp, 2020a,b). Adaptation to , prominent frequencies in the preceding acoustic context results in neural contrast when those frequencies change upon introduction of the target sound, producing a neural (and ultimately perceptual) shift. The neural mechanisms underlying TCEs are less clear, but two candidates have been suggested: cortical entrainment to modulations in the amplitude envelope (Bosker and Ghitza, 2018) or evoked responses to rapid increases in speech amplitude, particularly at modulation onset ("acoustic edges") (Oganian and Chang, 2019; Oganian *et al.*, 2023). When the rate of modulations or their onsets change across context and target stimuli, this similarly produces a contrastive shift where a larger change in rate is perceived, resulting in the TCE.

Many investigations of SCEs and TCEs share the common practice of taking one context stimulus (here, for simplicity, a sentence) and generating two renditions of it for use in the experiment. For SCEs, this includes a sentence with lower frequencies emphasized (to promote perception of higher frequencies in the target) and that same sentence but with higher frequencies emphasized (to promote perception of lower frequencies in the target). For TCEs, this includes a sentence resynthesized at a slower speaking rate (to promote perception of shorter durations / faster rate in the target) and that same sentence but resynthesized at a faster speaking rate (to promote perception of longer durations / slower rate in the target). This approach affords great experimental control by holding all variables constant except the one of primary interest, but at the same time, such control might be a departure from the extreme acoustic variability in everyday listening. Measuring the resiliency of acoustic context effects to different types of stimulus variability is an important step toward understanding the degrees to which they contribute to everyday speech perception. Here, we review the consequences of two types of stimulus variability on acoustic context effects in speech perception: the context sentences themselves and the talkers who spoke them.

The influence of context sentence variability on acoustic context effects in speech perception has not received much attention. Some studies have reported context effects following a wide variety of context sentences (e.g., Bosker *et al.*, 2020a,b), but without a comparison condition where one talker speaks only one sentence, the consequences of this variability were not clear. Assgari and Stilp (2015) tested the influence of context sentence variability on SCEs directly. In that study, on each trial, listeners were presented with a context sentence followed by a target vowel that listeners categorized as either /ɪ/ or /ɛ/. The context sentence was processed by one of two filters: the first spanned 100–400 Hz to emphasize lower F_1 frequencies (to encourage perception of /ɛ/), or the second spanned 550–850 Hz to emphasize higher F_1 frequencies (to encourage perception of /ɪ/). SCEs in vowel identification were measured following filtered renditions of a single context sentence (One Talker/One Sentence condition) or filtered renditions of a new sentence on each trial, all spoken by the same talker (One Talker/200 Sentences condition). SCEs in these conditions were equivalent. The explanation for this finding likely links back to the underlying mechanism of neural adaptation. While the long-term average spectrum for each sentence was clearly not identical from trial to trial, it was sufficiently similar in terms of the relevant frequencies being present and thus amplified to create spectral peaks (particularly in the last 500–1000 ms of the context sentence) (Stilp and Assgari, 2019, 2021; Shorey and Stilp, 2023). These spectral peaks promoted neural adaptation, producing the SCE.

Effects of sentence variability on TCEs have not previously been studied. The consequences of this variability are expected to be far more significant for TCEs than the null effects reported for SCEs. Whether TCEs are driven by amplitude modulations in a sentence or the modulation onsets, these differ substantially from one sentence to the next, even when sentences are spoken by the same talker. This is particularly true near sentence offset, which bears substantially greater influence on the resulting TCE than does earlier in the context sentence (e.g., Summerfield, 1981; Kidd, 1989; Reinisch *et al.*, 2011; Heffner *et al.*, 2013). Thus, context sentence variability would be predicted to result in diminished TCEs relative to a one-sentence baseline.

Talker variability is widely documented to challenge speech perception. When compared to listening to a single talker, speech perception is slower and/or less accurate when multiple talkers are presented across trials within a testing block (e.g., Magnuson and Nusbaum, 2007; Mullennix *et al.*, 1989; Sommers *et al.*, 1994). The relationship between SCEs and talker variability has been examined previously, but mostly in limited cases. Watkins (1991) presented context sentences spoken by one talker and target words spoken by another talker. The context sentences were filtered by the difference in spectral envelopes of /ɪ/ minus /ɛ/ or its inverse, while target words perceptually varied along an /ɪtʃ/ to /ɛtʃ/ continuum. Lotto and Kluender (1998) also used different talkers across context (/ɑl/ or /ɑr/) and target (/dɑ/ or /gɑ/). In each of these studies, SCEs were observed despite listeners hearing different talkers across context and target. Laing *et al.* (2012) mimicked the context sentences used by Ladefoged and Broadbent (1957), but manipulated F_1 and F_3 of a single male talker to create four different "talkers." Results indicated that talker information itself did not induce SCEs. Importantly, the approach by Laing *et al.* (2012) differed from Watkins (1991) and Lotto and Kluender (1998) as talker variability was not manipulated because listeners heard the same talker on each trial. Subsequently, Assgari and Stilp (2015) provided the first full test of how talker variability affected SCEs. In addition to the One Talker/One Sentence and One Talker/200 Sentences conditions detailed above, listeners also completed the vowel categorization task when a different context sentence was spoken by a different talker on each trial

(200 Talkers/200 Sentences condition). Compared with the One-Talker conditions, SCE magnitudes were diminished amidst concurrent talker and sentence variability. Continuing this line of research, Assgari *et al.* (2019) concluded that this result was driven by variability in different talkers' fundamental frequencies (*f*0s). When mean f0s were highly variable across context sentences, SCEs were significantly smaller compared to when context sentences had low variability in their mean *f*0s. Variability in talker *f*0 has been suggested to impact the harmonic resolution of spectral peaks added to context sentences, making neural adaptation to them less efficacious at producing SCEs relative to more consistent talker *f*0s (Mills *et al.*, 2022).

Effects of talker variability have been considered only narrowly for TCEs. Similar to the approaches of Watkins (1991) and Lotto and Kluender (1998) when investigating SCEs, changes in talker identity during the trial still yielded TCEs (Sawusch and Newman, 2000; Newman and Sawusch, 2009; Kawahara *et al.*, 2022; but see Diehl *et al.*, 1980). These were restricted cases, testing only two talkers reading the same sentence. Testing the resiliency of TCEs to talker variability also raises the question of perceptual sensitivity to variability in speaking rate. The speaking rates of a wide variety of talkers will vary considerably more than the rates of two talkers. This variability in rate would be expected to decrease the consistency and efficacy with which amplitude modulations/modulation onsets produce TCEs. Again, using the experimental conditions of Assgari and Stilp (2015) as a framework, speaking rate would not vary at all for a single talker speaking one sentence (a single slow rate and a single fast rate) and would vary somewhat for a single talker speaking 200 sentences (a narrow distribution of slow speaking rates, a narrow distribution of fast speaking rates). However, speaking rate would vary far more for 200 talkers speaking 200 sentences (a wider distribution of slow speaking rates, a wider distribution of fast speaking rates). In that condition, not only would the sentence variability be predicted to diminish TCE magnitudes as described above, but this talker (and concomitant rate) variability would be predicted to have a compounding effect that diminishes TCE magnitudes even further. Taken together, SCEs and TCEs are predicted to differ in their resiliency to stimulus variability in context sentences, with the latter particularly susceptible to compounding effects.

Finally, although TCE magnitudes indicate the size of the shift in perception, they do not necessarily reflect task difficulty. In the psychoacoustic tradition, psychometric function slopes are used as an indicator of task difficulty: steeper slopes indicate an easier task, and shallower slopes are indicative of greater difficulty. In speech categorization tasks, these slopes have been interpreted as reflecting the efficiency with which listeners use acoustic cues to categorize target stimuli (e.g., Winn *et al.*, 2016) and/or the uncertainty with which listeners make their categorization judgments (e.g., Clayards *et al.*, 2008). In acoustic context effects experiments, Assgari *et al.* (2019) assessed the roles of multiple sources of contextual variability on both SCE magnitudes and psychometric function slopes. Their second experiment assessed mixed talkers (men and women) while mean *f*0 variability was either low or high. They found significantly shallower slopes in the high variability condition relative to the low variability conditions (negative target-by-variability interaction). This indicated that listeners were more definitive in their categorization of the target sounds in the low variability condition. However, it should be noted that slopes were not significantly different in their first experiment, which was of similar design, but talker gender was blocked for each variability condition. As such, psychometric function slopes provide valuable information regarding task performance beyond the shift in perception due to the acoustic contrast effect.

The present study directly tests the roles of talker, sentence, and acoustic variability on TCEs in speech categorization. Across all experiments, the same context sentences and conditions (One Talker/One Sentence, One Talker/ 200 Sentences, and 200 Talkers/ 200 Sentences) from Assgari and Stilp (2015) were used, but speaking rates were manipulated to create fast and slow sentences. On each trial, participants heard one context sentence (either fast or slow) followed by one of ten targets that perceptually varied from "deer" to "tier." Participants then responded by categorizing the target word as either "deer" or "tier."

The overarching goal was to study the effects of contextual variability on TCEs, using the framework set forth by Assgari and Stilp (2015). In Experiment 1, sentences were set to either 50% or 150% of their original duration to create fast and slow speaking rates, respectively. Variability in slow rates differed considerably across conditions; the same was true for fast rates. Further restriction of extraneous variability was the goal of Experiment 2. Here, speaking rates were matched for fast and slow sentences across the One Talker/200 Sentences and 200 Talkers/200 Sentences conditions. Finally, speaking rate variability was extremely limited in Experiment 3; all three conditions used the same speaking rate for all fast sentences and a single speaking rate for all slow sentences.

For Experiment 1, relative to a One Talker/One Sentence baseline, TCE magnitudes were predicted to decrease upon the introduction of sentence variability (One Talker/200 Sentences condition), and decrease further when sentence variability is combined with talker variability (200 Talkers/200 Sentences condition). Due to the increased variability in both the number of talkers and number of sentences, psychometric function slopes are predicted to be shallower in 200 Talkers/200 Sentences condition relative to other conditions.

II. EXPERIMENT 1

A. Methods

1. Participants

Twenty undergraduate students participated in exchange for course credit. All self-reported being native English speakers with normal hearing.

2. Stimuli

Context Sentences. One Talker/One Sentence. This stimulus was a recording of an adult man (C.E.S.) reading the sentence, "This time, I want you to click on the word" (duration = 2293 ms, rate = 4.36 syllables/s). This stimulus was chosen as it produced TCEs in the categorization of the target used in a previous experiment (Stilp, 2019).

One Talker/200 Sentences. These stimuli were recordings of an adult man reading 200 different sentences (mean duration = 1739 ms, mean rate = 4.13 syllables/s) selected from the Hearing in Noise Test corpus (Nilsson *et al.*, 1994). This was a different adult man from the One Talker/One Sentence condition. Further, these are the same stimuli used in the One Talker/200 Sentences condition of Assgari and Stilp (2015).

200 Talkers/200 Sentences. These stimuli were recordings from 138 men and 62 women speaking 200 different sentences (mean duration = 2248 ms, mean rate = 4.98 syllables/s) selected from the Texas Instruments/Massachusetts Institute of Technology corpus (TIMIT) (Garofolo *et al.*, 1990). These are the same stimuli used in the 200 Talkers/200 Sentences condition of Assgari and Stilp (2015).

The speaking rates for all contexts were manipulated in Praat (Boersma and Weenink, 2021) using the time-domain pitch synchronous overlap and add (TD-PSOLA) algorithm (Moulines and Charpentier, 1990). In all three conditions, duration was manipulated to create fast sentences (at 50% of the original duration) and slow sentences (at 150% of the original duration), which consequently altered speaking rate (Fig. 1). In the One Talker/One Sentence condition, a fast and a slow version were created. For the One Talker/200 Sentences and 200 Talkers/200 Sentences conditions, half were made fast and half were made slow. Sentences were assigned to experimental conditions in the following ways. For the sentences that received low-F_1 amplification in Assgari and Stilp (2015), those same sentences (without any low-F_1 amplification) were set to slow speaking rates here. For the sentences that received high-F_1 amplification in Assgari and Stilp (2015), those same sentences (without any high-F_1 amplification) were set to fast speaking rates here. This avoids the potential confound of sentences being grouped together differently across studies, which would impede interpretations of patterns of results across studies (e.g., TCEs exhibiting different sensitivity to sentence and/or talker variability than SCEs).

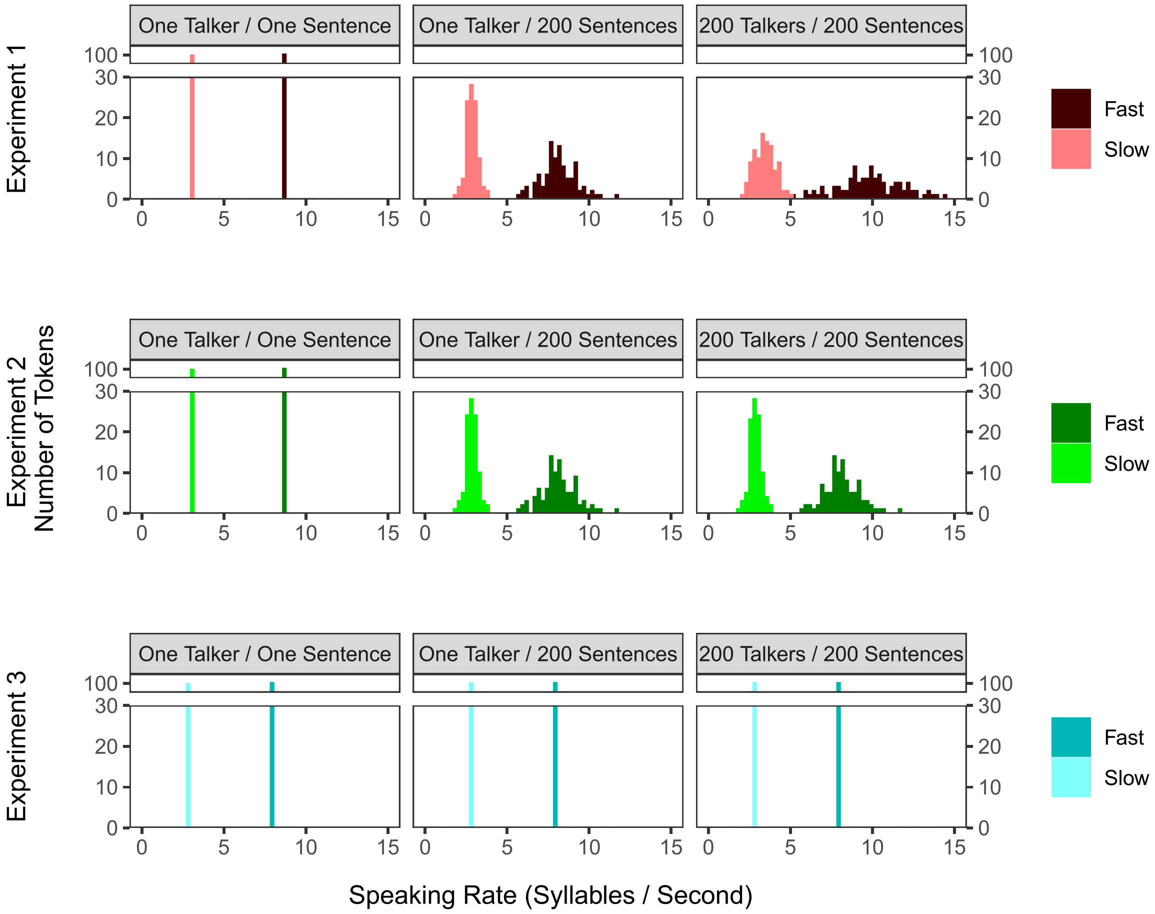

FIG. 1. (Color online) Histograms of the distribution of speaking rates for both slow (light colors) and fast (dark colors) sentences in all three conditions across experiments. All y axes include a break from 30 to 100 to aid readability. The left column shows the One Talker/One Sentence condition, the center column shows the One Talker/200 Sentences condition, and the right column shows the 200 Talkers/200 Sentences condition. The top row shows Experiment 1, the middle row shows Experiment 2, and the bottom row shows Experiment 3.

3. Targets

While Assgari and Stilp (2015) used vowels for target sounds, this study instead used two words that differed by their initial consonant sound, specifically in VOT. The same adult man from the One Talker/One Sentence condition was recorded saying the word “deer.” Using synthesis methods outlined by Winn (2020), a ten-step series of target words was created in Praat (Boersma and Weenink, 2021). This series perceptually varied from “deer” to “tier” by linearly increasing VOT from 21 ms at the “deer” endpoint to 82 ms at the “tier” endpoint. Secondary acoustic variations across the continuum included total duration (an overall lengthening from 488 to 512 ms across the continuum from “deer” endpoint to “tier” endpoint) and overall intensity (an overall decrease in intensity of 1.4 dB across the continuum as the initial stop consonant transitioned from being voiced to voiceless). Perception of these stimuli was confirmed to be influenced by TCEs in a previous study (Kahloon *et al.*, 2023). Experimental trials were created by concatenating a target word to either a slow-rate context sentence or a fast-rate context sentence, separated by a silent 50 ms interstimulus interval.

4. Procedure

After providing informed consent, the participant was seated in a sound-attenuating booth (Acoustic Systems, Inc., Austin, TX). Stimuli were D/A converted by RME HDSPe AIO sound cards (Audio AG, Haimhausen, Germany) on personal computers and passed through a programmable attenuator (TDT PA4, Tucker-Davis Technologies, Alachua, FL) and a headphone buffer (TDT HB6). Stimuli were presented over circumaural headphones (Beyerdynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY) at a mean presentation level of 70 dB sound pressure level. A custom script in MATLAB (Mathworks, Natick, MA) led the listener through the experiment, which was self-paced.

Participants first completed a practice block of 20 trials. Each trial presented a context sentence at its native speaking rate, followed by one of the two endpoints words of the “deer”-“tier” continuum. Sentences were taken from the AzBio corpus (Spahr *et al.*, 2012). At the end of each trial, a response window appeared with buttons labeled “deer” (top button) and “tier” (bottom button). Participants used the mouse to click the button corresponding to their response in two-alternative forced choice format. A performance criterion of $\geq$ 80% correct was required to continue; if participants did not meet this criterion, they were able to repeat the practice block up to two more times. All participants met this performance criterion.

The main experiment consisted of three blocks (One Talker/One Sentence, One Talker/200 Sentences, and 200 Talkers/200 Sentences), each with 200 trials. The One Talker/One Sentence block included 100 presentations of the fast and slow versions of the single context sentence, respectively. Each of the 200 Sentences conditions included 100 fast context sentences and 100 slow context sentences. Stimuli were presented in random order, and the blocks were tested in counterbalanced orders. Participants were able to take a small break between each block if they chose. The entire experiment took approximately 50 min.

B. Results

Data were analyzed via generalized linear mixed-effects modeling in R using the lme4 package (Bates *et al.*, 2014). The dependent variable was participants’ responses (0 = “deer,” 1 = “tier”). Responses were predicted based on the following predictors: target, condition, and speaking rate. The ten different target sounds on the “deer” to “tier” continuum were coded as 1–10, then mean centered. Condition refers to the three experimental blocks: One Talker/One Sentence, One Talker/200 Sentences, and 200 Talkers/200 Sentences. This was dummy coded with One Talker/One Sentence as the default level. Finally, speaking rate refers to the slow and fast context sentences presented to listeners. This was sum coded so that slow was -0.5 and fast was $+0.5$.

The model building process began by creating a base mixed-effect model, which included random intercepts for participants along with fixed main effects for target, condition, speaking rate, and the interactions between each. Using this model as a starting point, additional models were created that added one random slope at a time for each of the predictor variables. This new model would then be tested against the previous model using a chi-squared goodness of fit test. If the added term resulted in a significantly improved model fit, this new model was retained. This process was repeated until reaching the maximal random effects structure that also allowed the model to converge. The final model included fixed effects for target, condition, and speaking rate as well as their interactions, as well as random slopes for target and condition and random intercepts for participants.

Results are presented in Fig. 2 and listed in Table I. Responses showed a significant negative intercept ($z = -5.58$, $p < 0.001$), indicating that in the default condition (One Talker/One Sentence), participants responded “tier” less often (44.2% of the time) than they responded “deer.” As expected, participants responded “tier” more often as the target moved along the continuum from “deer” to “tier” ($\hat{\beta} = 2.48$ reflecting the increase in log-odds of responding “tier” with each step along the target continuum toward the “tier” endpoint; $z = 18.22$, $p < 0.001$). Also, as expected, “tier” responses increased as the context sentence speaking rate was changed from slow to fast, consistent with the predicted direction of TCEs (a mean shift of 9.5% “tier” responses; $z = 11.21$, $p < 0.001$). Relative to the One Talker/One Sentence condition, participants responded “tier” significantly more often in the One Talker/200 condition (47.6% of responses; $z = 3.06$, $p < 0.01$), as well as in the 200 Talkers/200 Sentences condition (50.7%; $z = 7.16$, $p < 0.001$). Interactions between target and condition highlight the differences in psychometric function slopes across

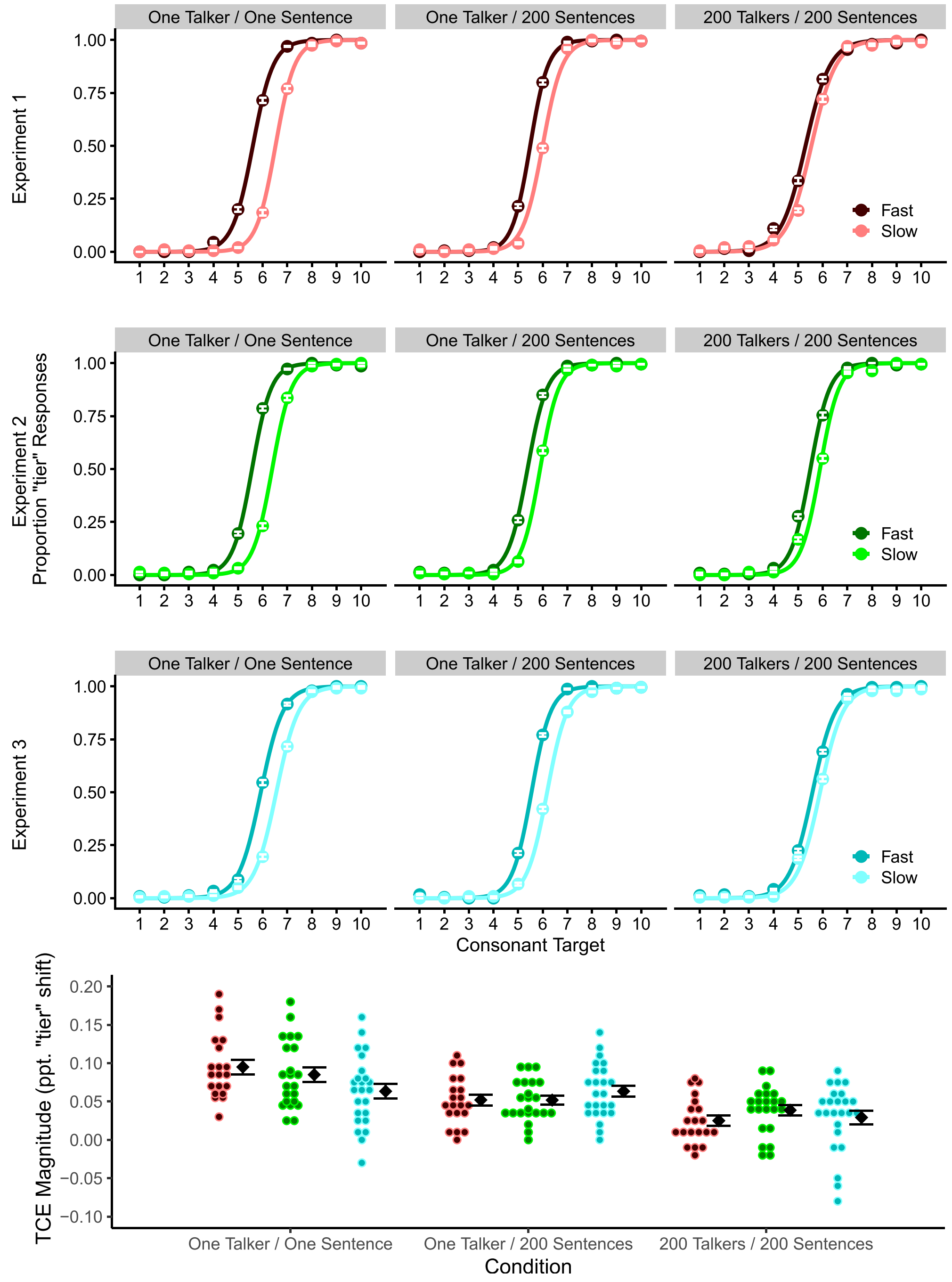

FIG. 2. (Color online) Rows one through three display results from the generalized linear mixed-effects model analysis of responses from each experiment (Experiment 1 = red colors in row 1, Experiment 2 = green colors in row 2, Experiment 3 = blue colors in row 3). The *y* axis represents the proportion of "tier" responses provided by participants. The *x* axis shows the continuum of target words, anchored by "deer" at 1 and "tier" at 10. Functions are shown for slow sentences (light) and fast sentences (dark). Each dot represents the mean number of "tier" responses for each target word. Error bars display standard error. TCE magnitudes are represented by the gap between the functions. The bottom row displays individual TCEs across each condition (by column) and experiment (by color). Means (diamonds) and standard error are presented to the right of each condition.

conditions. Compared to the default condition (One Talker/One Sentence), psychometric function slopes were significantly shallower in the 200 Talkers/200 Sentences condition ($\hat{\beta} = 2.00$, $z = -4.21$, $p < 0.001$) but were comparable in the One Talker/200 Sentences condition ($\hat{\beta} = 2.61$, $z = 0.88$, $p = 0.38$). Most importantly, the significant rate by condition interactions indicate that TCE magnitudes decreased from the One Talker/One Sentence condition to the One Talker/200 Sentences condition (mean shift of "tier" responses = 5.2%; $z = -4.10$, $p < 0.001$) and also from the One Talker/One Sentence condition to the 200 Talkers/200 Sentences condition (mean shift of "tier" responses = 2.5%; $z = -7.53$, $p < 0.001$).

By setting One Talker/One Sentence as the default level of condition, the model did not make paired comparisons between the other two conditions (One Talker/200

TABLE I. Results from the generalized linear mixed-effects model analysis of responses in Experiment 1. As described in the text, target was coded 1–10 then mean centered. Condition was dummy coded with the One Talker/One Sentence condition as the default level. Speaking rate was sum coded with slow as –0.5 and fast as +0.5.

	$\hat{\beta}$	*SE*	*Z*	*p*
Intercept	−1.430	0.256	−5.583	< 0.001
Target	2.482	0.136	18.221	< 0.001
Rate (Slow)	2.379	0.212	11.211	< 0.001
One Talker/200 Sentences	0.828	0.270	3.061	0.002
200 Talkers/200 Sentences	1.595	0.223	7.164	< 0.001
Target × Rate	−0.163	0.170	−0.961	0.337
Target × One Talker/200 Sentences	0.124	0.141	0.878	0.380
Target × 200 Talkers/200 Sentences	−0.485	0.115	−4.211	< 0.001
Rate × One Talker/200 Sentences	−1.117	0.273	−4.098	< 0.001
Rate × 200 Talkers/200 Sentences	−1.904	0.253	−7.532	< 0.001
Target × Rate × One Talker/200 Sentences	0.414	0.258	1.607	0.108
Target × Rate × 200 Talkers/200 Sentences	0.164	0.206	0.797	0.425

Sentences and 200 Talkers/200 Sentences). To make these comparisons, the model was run again with the One Talker/200 Sentences condition as the default level. Results are provided in Supplementary Table I. Relative to the One Talker/200 Sentences condition, psychometric function slopes became significantly shallower in the 200 Talkers/200 Sentences condition ($z = –4.95$, $p < 0.001$) and TCE magnitudes significantly decreased between these two conditions as well ($z = –3.58$, $p < 0.001$).

C. Discussion

Experiment 1 used fast and slow context sentences to assess target word perception amid different levels of talker and sentence variability. TCE magnitudes decreased when a different sentence was heard on each trial (One Talker/One Sentence vs One Talker/200 Sentences) and decreased again when a different talker spoke a different sentence on each trial (One Talker/200 Sentences vs 200 Talkers/200 Sentences). This pattern of results is similar but not identical to those of Assgari and Stilp (2015) when the same context sentences were used. In that study, SCE magnitudes were comparable when a different sentence was heard on each trial (that is, One Talker/One Sentence vs One Talker/200 Sentences) but did significantly decrease when a different talker said a different sentence on each trial (One Talker/200 Sentences vs 200 Talkers/200 Sentences). Further comparisons between this study and Assgari and Stilp (2015) are provided in the General Discussion.

Interestingly, psychometric function slopes in each condition did not follow the same pattern of results as TCE magnitudes. The slopes of psychometric functions were comparable in both conditions featuring one talker (One Talker/One Sentence and One Talker/200 Sentences). Trial-by-trial variability in context sentences did not appear to make the task more difficult (i.e., produce shallower slopes) but did result in smaller TCEs. This differs from the pattern of results in multiple-talker conditions, in which psychometric slopes and TCEs changed together. Slopes were significantly shallower and TCEs were significantly smaller in the 200 Talkers/200 Sentences condition. As such, task difficulty does not appear to be clearly intertwined with TCE magnitudes.

Of note, the conditions in this experiment were constructed to vary in the number of talkers and the number of sentences spoken. However, there was also variation in the speaking rates of both fast [Brown–Forsythe test: $F(1, 198) = 26.60$, $p < 0.001$] and slow [Brown–Forsythe test: $F(1, 198) = 33.70$, $p < 0.001$] context sentences across conditions (see Fig. 1). That is, the distribution of speaking rates (in syllables per second) for the One Talker/200 Sentences condition is much narrower than that of the 200 Talkers/200 Sentences condition. While this might be expected, given a much wider range of speakers, it precludes attributing smaller TCE magnitudes in the 200 Talkers/200 Sentences condition to only talker and/or sentence variability. Because the distribution of speaking rates varied between conditions, these conditions meaningfully differed from the outset. As such, making these experimental conditions more similar to each other provides the opportunity to study effects of talker variability when rate variability is matched. Thus, more accurate assessments of how talker and/or sentence variability affect TCE magnitudes can be made. Experiment 2 addressed this matter by changing the speaking rates in the 200 Talkers/200 Sentences condition to match the speaking rates in One Talker/200 Sentences. Both TCE magnitudes and psychometric function slopes were predicted to be equal across these conditions.

III. EXPERIMENT 2

A. Methods

1. Participants

Twenty-two undergraduate students participated in exchange for course credit. All self-reported being native English speakers with normal hearing. None participated in Experiment 1.

2. Stimuli

Context Sentences. The same original context sentences from Experiment 1 were used again in Experiment 2. However, while the same sentences were used across experiments, Experiment 2 included an additional manipulation. As shown in Fig. 1, the distributions of speaking rates were markedly narrower in the One Talker/200 Sentences condition than in the 200 Talkers/200 Sentences condition in Experiment 1. To address this, the speaking rates of the context sentences in the 200 Talkers/200 Sentences condition of Experiment 2 were altered to match those of the One Talker/200 Sentences condition (see Fig. 1).

All sentences were first organized from slowest to fastest speaking rates. As an example, consider the leftmost sentence in the slow-rate One Talker/200 Sentences condition from Experiment 1 (top center panel of Fig. 1). This sentence has the slowest speaking rate among all context sentences in the slow-rate One Talker/200 Sentences condition. This was used as a reference point for the leftmost sentence in the slow-rate 200 Talkers/200 Sentences condition (top right panel of Fig. 1) – the sentence with the slowest speaking rate among all context sentences in the slow-rate 200 Talkers/200 Sentences condition from Experiment 1. These sentences were paired together. The speaking rate for the sentence from 200 Talkers/200 Sentences was divided by the speaking rate for the sentence from One Talker/200 Sentences. Then, using the TD-PSOLA algorithm in Praat, duration of the sentence from the 200 Talkers/200 Sentences condition was multiplied by that value. This provided a new duration for that sentence, which now matched the speaking rate of its partner sentence from the One Talker/200 Sentences condition. This process continued with each sentence for the slow-rate context sentences, then repeated for each fast-rate context sentence. As such, all speaking rates of the context sentences were matched across these two conditions. The distributions of speaking rates in the One Talker/200 Sentences conditions were the same across Experiments 1 and 2 (center column of Fig. 1). However, the distributions of speaking rates in the 200 Talkers/200 Sentences conditions changed across Experiments 1 and 2 (right column of Fig. 1).

Targets. The same “deer” to “tier” target continuum from Experiment 1 was used in Experiment 2.

3. Procedure

The same procedure from Experiment 1 was used in Experiment 2. All 22 participants met the criteria for inclusion. All findings hold when using the full sample of 22 participants as well as the first 20 participants (matching the sample size of Experiment 1).

B. Results

The generalized linear mixed-effects model building process as described in Experiment 1 was used. The final model included fixed effects for target, condition, and speaking rate as well as their interactions (coded in the same manner as described for Experiment 1), as well as random slopes for target, condition, and speaking rate, and random intercepts for participants.

Results are illustrated in Fig. 2 and listed in Table II. Participants responded with “tier” less often in the default condition (45.4% of the time in the One Talker/One Sentence), as indicated by the significant negative intercept ($z = –5.60$, $p < 0.001$). As expected, participants responded “tier” more often as the target moved from “deer” to “tier” ($\hat{\beta} = 2.55$ reflecting the increase in log-odds of responding “tier” with each step along the target continuum toward the “tier” endpoint; $z = 14.38$, $p < 0.001$). Likewise, “tier” responses also increased with changes in rate from slow to fast, which matched the predicted direction of TCEs (mean shift of 8.50% “tier” responses; $z = 10.06$, $p < 0.001$). When compared to the default condition of One Talker/One Sentence, participants responded “tier” significantly more often in the One Talker/200 Sentences (48.8% of responses; $z = 5.57$, $p < 0.01$), as well as in the 200 Talkers/200 Sentences condition (48.5%; $z = 3.79$, $p < 0.01$). Unlike Experiment 1, no differences in psychometric function slopes were found across conditions ($\hat{\beta} = 2.56 – 2.58$; target by condition interactions: $z < 0.19$, $p > 0.85$). However, significant rate by condition interactions were found, indicating that TCE magnitudes decreased from the One Talker/One Sentence condition to the One Talker/200 Sentence condition (mean shift of 5.18% “tier” responses; $z = –3.18$, $p < 0.01$), and also from the One Talker/One Sentence condition to the 200 Talker/200 Sentence condition (mean shift of 3.86% “tier” responses; $z = –4.64$, $p < 0.001$).

TABLE II. Results from the generalized linear mixed-effects model analysis of responses in Experiment 2. As described in the text, target was coded 1–10 then mean centered. Condition was dummy coded with the One Talker/One Sentence condition as the default level. Speaking rate was sum coded with slow as –0.5 and fast as +0.5.

	$\hat{\beta}$	*SE*	*Z*	*p*
Intercept	−1.213	0.216	−5.601	< 0.001
Target	2.554	0.178	14.380	< 0.001
Rate (Slow)	2.083	0.207	10.056	< 0.001
One Talker/200 Sentences	0.840	0.151	5.570	< 0.001
200 Talkers/200 Sentences	0.718	0.189	3.791	< 0.001
Target × Rate	0.019	0.160	0.121	0.903
Target × One Talker/200 Sentences	0.022	0.120	0.188	0.851
Target × 200 Talkers/200 Sentences	0.008	0.119	0.068	0.946
Rate × One Talker/200 Sentences	−0.746	0.235	−3.182	0.001
Rate × 200 Talkers/200 Sentences	−1.075	0.232	−4.639	< 0.001
Target × Rate × One Talker/ 200 Sentences	−0.154	0.224	−0.687	0.492
Target × Rate × 200 Talkers/ 200 Sentences	−0.072	0.215	−0.337	0.736

As in Experiment 1, the One Talker/One Sentence condition was set as the default, so this model was not able to make paired comparisons between One Talker/200 Sentences and 200 Talkers/200 Sentences. To do so, the model was run again with the One Talker/200 Sentences condition set as the default level. Results are presented in Supplementary Table II. Neither TCE magnitudes ($z = –1.53$, $p = 0.13$) nor psychometric slopes ($z = –0.12$, $p = 0.90$) were significantly different between these two conditions.

1. Across-experiment analyses

Generalized linear mixed-effects modeling was employed to examine whether and how responses differed across experiments. The same model building process was once again utilized. In addition to the same fixed effects as before, Experiment (sum coded with Experiment 1 as −0.5 and Experiment 2 as +0.5) was also included. The final model included fixed effects for target, condition, speaking rate, and experiment as well as their interactions, random slopes for target, condition, and speaking rate, and random intercepts for participants. The full results of these analyses are available in Supplementary Tables III–V; here, changes in psychometric function slopes and TCE magnitudes across experiments are of primary interest.

By restricting the variability in speaking rates in the 200 Talkers/200 Sentences condition, TCE magnitudes marginally increased in Experiment 2 (Supplementary Table V, Rate × Experiment: $z = 1.86$, $p = 0.063$); TCE magnitudes in the other conditions did not change across experiments (Supplementary Tables III and IV, Rate × Experiment: $z > -1.01$, $p > 0.315$). When comparing TCE magnitudes in the One Talker/200 Sentences and 200 Talkers/200 Sentences conditions, despite significantly differing from each other in Experiment 1 and not differing from each other in Experiment 2, these patterns of changes in TCEs did not significantly differ across experiments (Supplementary Table V, Rate × One Talker/200 Sentences × Experiment: $z = -1.44$, $p = 0.150$). On the other hand, the sharp decrease in TCE magnitudes across the One Talker/One Sentence and 200 Talkers/200 Sentences conditions in Experiment 1 was lessened in Experiment 2 (Supplementary Table V, Rate × One Talker/One Sentence × Experiment: $z = -2.18$, $p = 0.030$).

By restricting the variability in speaking rates in the 200 Talkers/200 Sentences condition, psychometric slopes in the 200 Talker/200 Sentences condition became steeper in Experiment 2 (Supplementary Table V, Target × Experiment: $z = 2.25$, $p = 0.025$); slopes in the other conditions did not change across experiments (Supplementary Tables III and IV, Target × Experiment: $z > -0.358$, $p > 0.720$). Psychometric function slopes in Experiment 1 were shallower in the 200 Talkers/200 Sentences condition than in other conditions but were equal across all conditions in Experiment 2. These patterns of slopes significantly differed across experiments (Supplementary Table V: Target × One Talker/200 Sentences × Experiment: $z = -3.32$, $p = 0.001$; Target × One Talker/One Sentence × Experiment: $z = -2.54$, $p = 0.011$).

C. Discussion

Experiment 2 was designed to more tightly control for contextual variability compared to Experiment 1, in which conditions with 200 talkers contained fundamentally different speaking rates (compare the top row of Fig. 1 to the center row of Fig. 1). After adjusting speaking rates in the 200 Talkers/200 Sentences condition to match those in the One Talker/200 Sentences condition, TCEs across these conditions did not significantly differ. That is, presenting multiple talkers in a block did not significantly influence TCE magnitudes when speaking rate was better controlled. This differs from Experiment 1, where TCE magnitudes decreased with added sentence variability and again with added talker variability. Differences in TCE magnitudes across experiments were not significant but trended toward differing in the predicted direction. Larger TCE magnitudes in the 200 Talkers/200 Sentences condition were found in Experiment 2, but not to a statistically significant degree.

Furthermore, psychometric slopes did not differ across the conditions of interest (One Talker/200 Sentences and 200 Talkers/200 Sentences) in Experiment 2. Again, this differs from Experiment 1, where slopes were steeper in the One Talker/200 Sentences condition. While task difficulty might not directly be tied to TCE magnitudes, accounting for differences in speaking rate might mitigate some of this difference. Psychometric slopes in the 200 Talkers/200 Sentences condition were steeper in Experiment 2 than in Experiment 1. Again, in Experiment 2, there were no differences in psychometric slopes across One Talker/200 Sentences and 200 Talkers/200 Sentences. Part of the variability in adapting to multiple talkers was removed (or at least mitigated) in Experiment 2. Furthermore, the range of speaking rates in the 200 Talkers/200 Sentences condition was much greater in Experiment 1 compared to Experiment 2. By creating a matched and more compact distribution, participants were more decisive in their categorizations. This impacted results such that switching between 200 talkers was equal to that of one talker saying 200 sentences.

Results from Experiment 2 underscore that talker variability alone is not the driving force behind TCEs. Matching speaking rates across conditions in which multiple sentences were spoken resulted in no difference in TCE magnitude. Thus, acoustic variability (the distributions of speaking rates heard) influences perception separately from talker and/or sentence variability. These results follow work from Drown and Theodore (2020), which used a speeded word identification task to measure word recognition time amidst talker variability (hearing a single versus mixed talkers in a given block) and token variability (the mixed talkers being consistent/exhibiting low acoustic variability versus inconsistent/exhibiting high acoustic variability within a block). Their results indicated that both talker variability and token variability incur processing costs (i.e., slower response times). Importantly, confounding the two leads to additional costs beyond either source of variability alone. When stimuli are more highly structured, and thus less variable across multiple dimensions, processing costs were lessened. As such, when manipulating stimulus variability, lower-level acoustic variability must be considered as well (for similar results, see Kapadia *et al.*, 2023). Here, imparting more structure across the multiple-talker conditions resulted in fewer processing costs (i.e., steeper psychometric function slopes).

In Experiment 2, acoustic variability was matched across multiple conditions, nullifying differences in TCE magnitudes that were observed in Experiment 1. However, TCEs in the One Talker/One Sentence condition were again larger than those in other conditions. Furthermore, psychometric slopes were equal across conditions in Experiment 2. This raises the question of whether further restrictions in context sentence acoustic variability could eliminate all differences (in both TCEs and slopes) across conditions. To test this, Experiment 3 further decreased acoustic variability by setting one fast rate and one slow rate across all three conditions. Both TCE magnitudes and psychometric function slopes were predicted to be equal across all three conditions.

IV. EXPERIMENT 3

A. Methods

1. Participants

Twenty-four undergraduate students participated in exchange for course credit. All self-reported being native English speakers with normal hearing. None participated in Experiment 1 nor Experiment 2.

2. Stimuli

Context Sentences. The same context sentences from Experiments 1 and 2 were used in Experiment 3. Like Experiment 2, these sentences were manipulated to match speaking rates across conditions (see Fig. 1). All fast sentences were set to 8.0 syllables/s in all three conditions; similarly, all slow sentences were set to 2.67 syllables/s in all three conditions. These rates are similar to the speaking rates in the One Talker/One Sentence conditions from Experiments 1 and 2 (fast = 8.72 syllables/s, slow = 2.91 syllables/s), and firmly within the range of the speaking rates tested in other conditions of Experiment 2 (fast median = 8.16 syllables/s, slow median = 2.76 syllables/s). Speaking rates were manipulated using the TD-PSOLA algorithm using the process detailed in Experiment 2.

*Targ*ets. The same "deer" to "tier" target continuum from Experiments 1 and 2 was used in Experiment 3.

3. Procedure

The same procedure from Experiments 1 and 2 was used in Experiment 3. All 24 participants met the criteria for inclusion. All findings hold when using the full sample of 24 participants as well as the first 20 participants (matching the sample size of Experiment 1).

B. Results

The generalized linear mixed-effects model building process as described in Experiment 1 and 2 was used. The final model included fixed effects for target, condition (coded with One Talker/One Sentence as the default), and speaking rate (sum coded with slow as −0.5 and fast as +0.5) as well as their interactions, random slopes for target and speaking rate, and random intercepts for participants. Results are illustrated in Fig. 2 and listed in Table III.

TABLE III. Results from the generalized linear mixed-effects model analysis of responses in Experiment 3. As described in the text, target was coded 1–10, then mean centered. Condition was dummy coded with the One Talker/One Sentence condition as the default level. Speaking rate was sum coded with slow as –0.5 and fast as +0.5.

	$\hat{\beta}$	*SE*	*Z*	*p*
Intercept	−1.632	0.209	−7.798	< 0.001
Target	2.213	0.120	18.412	< 0.001
Rate (Slow)	1.386	0.205	6.774	< 0.001
One Talker/200 Sentences	0.753	0.114	6.633	< 0.001
200 Talkers/200 Sentences	1.128	0.106	10.631	< 0.001
Target × Rate	0.020	0.132	0.153	0.878
Target × One Talker/200 Sentences	0.277	0.102	2.725	0.006
Target × 200 Talkers/200 Sentences	−0.073	0.088	−0.829	0.407
Rate × One Talker/200 Sentences	0.145	0.227	0.641	0.522
Rate × 200 Talkers/200 Sentences	−0.725	0.212	−3.424	0.001
Target × Rate × One Talker/ 200 Sentences	0.114	0.203	0.561	0.575
Target × Rate × 200 Talkers/ 200 Sentences	−0.015	0.175	−0.084	0.933

The negative intercept in the One Talker/One Sentence condition (z = –7.80, $p < 0.001$), indicated that participants responded with "tier" less often (42.7% of the time), while the significant main effect of target indicates participants responded "tier" more often as the target moved from "deer" to "tier" ($\hat{\beta} = 2.21$ reflecting the increase in log-odds of responding "tier" with each step along the target continuum toward the "tier" endpoint; $z = 18.41$, $p < 0.001$). Similarly, as rate changed from slow to fast, "tier" responses also increased, matching the predicted direction of TCEs (mean shift of 6.33% "tier" responses; $z = 6.77$, $p < 0.001$). Compared to the One Talker/One Sentence condition, "tier" responses were provided more often in both the One Talker/200 Sentences condition (46.6% of responses; $z = 6.63$, $p < 0.01$) and the 200 Talkers/200 Sentences condition (48.0% of responses; $z = 10.63$, $p < 0.01$). Psychometric slopes differed between the One Talker/One Sentence condition and One Talker/200 Sentences condition, such that slopes became unexpectedly steeper with increased sentence variability ($\hat{\beta} = 2.49$; $z = 2.73$, $p < 0.01$), but no difference was observed between slopes in the One Talker/One Sentence and 200 Talkers/200 Sentences conditions ($\hat{\beta} = 2.14$; z = –0.83, $p = 0.41$). Finally, compared to the default condition of One Talker/One Sentence, TCEs were comparable in the One Talker/200 Sentences condition (mean shift of 6.33% "tier" responses; $z = 0.64$, $p = 0.52$) and decreased in the 200 Talkers/200 Sentences condition (mean shift of 2.92% "tier" responses; z = –3.42, $p < 0.01$).

To make comparisons across One Talker/200 Sentences and 200 Talkers/200 Sentences, the model was run again with the One Talker/200 Sentences condition set as the default level. Results are presented in Supplementary Table VI. Compared to the One Talker/200 Sentences condition, slopes were shallower (z = –3.55, $p < 0.01$) and TCE magnitudes were significantly smaller in the 200 Talkers/200 Sentences condition (z = –4.29, $p < 0.01$).

1. Across-experiment analyses

Results across Experiments 2 and 3 were examined with generalized linear mixed-effects modeling used to predict "tier" responses. The model building process mirrored that as described previously. The final model included fixed effects for target, condition, speaking rate (sum coded with slow as −0.5 and fast as +0.5), and experiment (sum coded

https://doi.org/10.1121/10.0025292

with Experiment 2 as −0.5 and Experiment 3 as +0.5) as well as their interactions, along with random slopes for target, condition, and speaking rate, with random intercepts for participants. Full results of this analysis are available in Supplementary Tables VII–IX. Once again, changes in TCE magnitudes and psychometric function slopes across experiments are of primary interest.

A significant rate by experiment interaction was found only when One Talker/One Sentence was used as the default condition ($z = -2.15$, $p = 0.03$), indicating smaller TCEs in this condition in Experiment 3. There was an increase in the difference in TCE magnitudes between One Talker/One Sentence and One Talker/200 Sentences across experiments ($z = 2.58$, $p < 0.01$). That is, TCE magnitudes decreased with additional sentence variability in Experiment 2, but not in Experiment 3. Across the One Talker/200 Sentences and 200 Talkers/200 Sentences, patterns of TCE magnitudes trended toward significance ($z = 1.76$, $p = 0.078$). This suggests that the difference in TCE magnitudes across One Talker/200 Sentences and 200 Talkers/200 Sentences grew slightly when moving from Experiment 2 to Experiment 3. Finally, patterns of TCE magnitudes between One Talker/One Sentence and 200 Talkers/200 Sentences were not significant ($z = -1.02$, $p = 0.31$), indicating no difference in patterns across experiments.

No significant target by experiment interactions were found. The only pattern of psychometric slopes that trended toward significance was that between One Talker/200 Sentences and 200 Talkers/200 Sentences ($z = 1.91$, $p = 0.056$), suggesting a marginally larger difference in slopes across these conditions in Experiment 3 than in Experiment 2 (where they did not significantly differ).

C. Discussion

Relative to previous experiments, Experiment 3 continued this process of limiting acoustic variability in context sentences, using a single fast speaking rate and a single slow speaking rate for all three conditions. TCE magnitudes and psychometric function slopes were predicted to be equal across all three conditions. Contrary to this prediction, TCE magnitudes were smaller in the 200 Talkers/200 Sentences condition than in other conditions. Even though fast and slow speaking rates were the same across conditions, TCE magnitudes were smaller when talkers changed across each trial. Also, contrary to the prediction, psychometric function slopes were unexpectedly significantly steeper in the One Talker/200 Sentences condition. Although speaking rates were matched across all three conditions, there still appears to be one or more additional sources of variability that are playing an important role in listeners' responses. This is discussed further in the General Discussion.

The TCE magnitude in the One Talker/One Sentence condition being smaller in Experiment 3 than in Experiment 2 is not wholly unexpected. Although the fast and slow rates were not the same across experiments, they were very similar. Fast sentences were spoken at a rate of 8.72 syllables/s in Experiments 1 and 2; this was set to 8.0 syllables/s in Experiment 3. Slow sentences were spoken at a rate of 2.91 syllables/s in Experiments 1 and 2; this was set to 2.67 syllables/s in Experiment 3. Thus, the range in rates across slow and fast sentences was smaller in Experiment 3 (5.33 syllables/s) than in Experiment 2 (5.81 syllables/s). TCE magnitudes change monotonically as a function of the differences between speaking rates (Pickett and Decker, 1960; Summerfield, 1981). Therefore, the smaller range between fast and slow speaking rates in Experiment 3 likely accounts for the decrease in TCE magnitudes in the One Talker/One Sentence condition.

V. GENERAL DISCUSSION

Speech perception is influenced by the surrounding acoustic context. Previous research has assessed how variability in preceding acoustic context affects SCEs. Yet, the effect of contextual variability on TCEs was unknown. The present study assessed the relationships between multiple sources of contextual variability–talker, sentence, and acoustic–and TCEs by varying the speaking rates, the number of talkers, and number of sentences across conditions.

In Experiment 1, listeners heard context sentences spoken at fast (50% of original duration) and slow (150% of original duration) rates, followed by one of ten target words that perceptually varied from "deer" to "tier." Both sentence and talker variability decreased TCE magnitudes, while only talker variability resulted in shallower slopes. Experiment 2 used the same context sentences, but further manipulated speaking rates in order to match rate distributions across One Talker/200 Sentences and 200 Talkers/200 Sentences. No difference in TCE magnitude was found between these conditions, and slopes were equal in all conditions. Finally, in Experiment 3, context sentences were set to the same fast rate and the same slow rate in all three conditions. Talker variability, but not sentence variability, decreased TCE magnitudes; slopes were steeper in the One Talker/200 Sentences condition relative to other conditions. Results highlight the intricate relationship that multiple sources of variability (acoustic, talker, and sentence, and potentially others, discussed below) have on TCEs in speech perception.

There is a long-standing practice of using highly controlled stimulus materials when studying acoustic context effects (i.e., two renditions of a single context sentence, either amplified in different frequency regions to elicit a SCE or resynthesized at different speaking rates to elicit a TCE). However, this high experimental control does not necessarily reflect the extreme acoustic variability inherent to everyday listening conditions. Studying the resiliency of acoustic context effects to different types of stimulus variability may inform the degrees to which these processes contribute to everyday perception. In Assgari and Stilp (2015), SCEs were resilient to variability in the contents of context sentences but not to variability in who spoke them. In the present study, TCEs were susceptible to both sentence variability and talker variability. From these results, one might be tempted to conclude that SCEs are more resilient to the pervasive variability in everyday listening conditions than TCEs are, but a degree of caution is warranted. Only two types of stimulus variability were explored here and in Assgari and Stilp (2015), and only for two speech sound contrasts (/ɪ/-/ɛ/, /d/-/t/). Extending the present focus on stimulus variability to other stimuli and to more naturalistic experimental paradigms (where context sentences already have the desired acoustic properties; e.g., Reinisch, 2016; Stilp and Assgari, 2019, 2021) will shed further light on how these acoustic context effects shape everyday speech perception.

Despite using the same context sentences to study acoustic contrast effects in speech perception, results patterned differently from Assgari and Stilp (2015). There are several reasons for this. First and foremost, it bears reminding that different types of contrast effects were measured across studies: Assgari and Stilp (2015) analyzed SCEs and the present study analyzed TCEs. Second, each effect is proposed to be subserved by different neural mechanisms: SCEs by neural adaptation (Stilp, 2020a,b) and TCEs by either entrainment to modulations in the amplitude envelope of speech (Bosker and Ghitza, 2018) or evoked responses to rapid increases in speech amplitude (Oganian and Chang, 2019; Oganian *et al.*, 2023). Third, while studies used the same context sentences, the target stimuli differed. Assgari and Stilp (2015) presented isolated target vowels /ɪ/-/ɛ/ whereas here, the target stimuli were initial consonants in the words "deer" and "tier." Finally, SCEs are highly sensitive to variability in the mean *f*0 of context sentences: talkers with low variability in their mean *f*0s from trial to trial produced larger SCEs than talkers with high variability in mean *f*0 from trial to trial (Assgari *et al.*, 2019). The same source of stimulus variability is unlikely to explain the present results. While the One Talker/One Sentence condition had zero variability in mean *f*0 from trial to trial (because it was the same token spoken at different rates), there was minimal variability in mean *f*0 for the One Talker/200 Sentences condition, yet TCEs were significantly smaller in this condition in Experiments 1 and 2. This is suggestive of a different type of stimulus variability being responsible for diminishing TCE magnitudes across conditions, most likely one tied to the proposed neural mechanisms underlying TCEs. By presenting a different sentence on each trial, there was variability in the amplitude envelope of each sentence (as suggested by Bosker and Ghitza, 2018, to underlie TCEs) as well as the timing and frequency of rapid increases of signal amplitude (as suggested by Oganian *et al.*, 2023, to underlie TCEs). Targeted experimentation (akin to the experiments reported by Assgari *et al.*, 2019) is needed to identify the specific cause of variation in TCE magnitudes in the present results.

While talker variability, sentence variability, and acoustic variability all meaningfully shape perception, not all sources of variability are equally consequential. For example, in assessing acoustic context effects in vowel categorization tasks, variability in mean *f*0 for context sentences alters SCEs (Assgari *et al.*, 2019), but variability in mean F_1 or mean F_3 does not (Mills *et al.*, 2022). Furthermore, speaking rate variability impedes word recognition performance, but amplitude variation does not (Sommers *et al.*, 1994). Here, speaking rates were matched across conditions in two experiments: One Talker/200 Sentences matched with 200 Talkers/200 Sentences in Experiment 2, and all three conditions matched in Experiment 3. However, these sentences still differed in other (potentially perceptually salient) properties. Notably, duration and the number of syllables were not equal across sentences. Even in Experiment 3, where all fast and all slow speaking rates were set to a single rate, duration varied freely. That is, even though sentences had the same number of syllables per second for each speaking rate, there were not the same number of syllables in each sentence. Given the proposed neural mechanisms underlying TCEs, consistency in sentence duration and/or the number of syllables might be expected to facilitate neural entrainment or evoked responses to modulation onsets, thus contributing to larger TCEs. Conversely, more variable sentence durations/number of syllables might impede this processing and diminish TCE magnitudes. To explore these possibilities, the mixed-effects models reported above were rerun with three additional fixed effects: a main effect of context sentence duration (mean-centered), the interaction between context sentence duration and speaking rate condition, and the three-way interaction between sentence duration, speaking rate condition, and experimental condition. Context sentence duration did not have a systematic effect on TCE magnitudes in any experiment (see Data Availability for analyses). Parallel analyses were conducted by assessing the fixed effects of the number of syllables per sentence (mean-centered), their interaction with speaking rate condition, and their interaction with speaking rate condition and experimental condition. Again, no systematic influence of the number of syllables in context sentences and TCEs was observed. Yet, it must be noted that these factors were not controlled in the present materials, which does not decisively rule out their contributions to TCE magnitude at large. Other sources of variability, such as lexical content or syntax, may also be important considerations.

Aside from these sentence-wide factors, there exist more local factors whose variability might also constrain TCE magnitudes. Here, we consider possibilities in the temporal order in which they would be encountered during the context sentence–target word trial structure. First, while entire sentences and their speaking rates were manipulated here, it has been repeatedly demonstrated that temporal context immediately preceding the target item exerts a greater influence on perception than context that is more temporally removed (Summerfield, 1981; Reinisch *et al.*, 2011; Heffner *et al.*, 2013; Reinisch, 2016). While Fig. 1 illustrates mean speaking rates calculated across the entire duration of context sentences, variability in rate information immediately preceding the target words might offer its own constraint on TCE magnitudes. Speaking rates (and their variability) near sentence offset were not controlled in the present experiments but merit further consideration in future research. Second, the temporal gap between items on a trial contributes materially to their perceptual grouping (e.g., Samuel and Pitt, 2003). Variability in the duration of the silent interval between context and target items can affect TCEs (Kim and Cho, 2013). This variability would particularly challenge neural entrainment that is proposed to code the speaking rate information that underlies TCEs (Bosker and Ghitza, 2018). Finally, while the critical speech sound distinction occurred in the initial position of target words here, subsequent context informs word recognition generally (Szostak and Pitt, 2013) and influences TCEs specifically (e.g., Miller and Liberman, 1979; Summerfield, 1981; Newman and Sawusch, 1996; Sawusch and Newman, 2000). Future research may compare how variability in these different temporal positions on a trial compare to the influence of variability in sentence contents and talkers on TCEs reported here.

Acoustic context effects have typically been investigated by measuring one context effect per listener sample. This approach cannot address the consistency with which listeners display these effects. Here, each listener sample completed three conditions in their respective experiment. This raises the question of how consistently listeners' TCE magnitudes changed when confronted with different degrees of (talker, sentence, speaking rate) variability. To examine this possibility, individual differences analyses were conducted on each experiment. TCE magnitudes were calculated as the difference in percent "tier" responses following slow-rate and fast-rate context sentences (as depicted at the bottom of Fig. 1; n. b., all following patterns of results are the same if TCEs are calculated as the number of stimulus steps separating category boundaries/50% points on psychometric functions). Within a given experiment, Pearson correlations were conducted on TCEs between pairs of conditions. All correlations failed to achieve statistical significance when correcting for multiple comparisons ($\alpha = 0.017$, or 0.05/3 comparisons within an experiment); only one comparison reached statistical significance with a more liberal α-level (without correcting for multiple comparisons; TCEs in the One Talker/200 Sentences and 200 Talkers/200 Sentences conditions in Experiment 3: $r = 0.42$, $p = 0.04$). Next, principal components analyses were conducted on TCE magnitudes for each experiment to assess potential consistency in context effects beyond the level of pairwise comparisons. If context effect magnitudes for a listener sample exhibit global coherence (beyond what could be seen via pairwise comparisons), the vast majority of covariance among TCE magnitudes would load onto the first (principal) component, with little covariance captured by additional components. Analyses were conducted in R using the prcomp function, which calculates the singular value decomposition directly on the data matrix of TCE magnitudes (n. b., equivalent results are observed by calculating eigenvalues on the covariance matrix of the data). No centering or rotation was performed. Across experiments, the amount of covariance loading onto the first component ranged from 53%–58%, with appreciable covariance loading onto the second component (24%–37%). Thus, neither local nor global analyses revealed clear patterns of individual differences that explained variation in TCE magnitudes in a given experiment. One possible contributing factor is the high degree of acoustic (as well as linguistic) variability in the sentence materials. Previous work by Heffner and Myers (2019) indicated that stimulus variability may limit the test/retest reliability of speaking rate normalization effects. Subsequent research on individual differences in temporal context effects would be well served by titrating this variability to find the point at which stimulus variability first becomes damaging to the consistency with which listeners complete these tasks.

As previously discussed and demonstrated here, there are important differences in how contextual variability affects SCEs and TCEs in speech perception. While research has examined the effect of contextual variability on both SCEs and TCEs, different target stimuli have been used to study each. This obscures whether the different patterns of results for SCEs and TCEs are primarily due to the acoustic domain under investigation or the specific target stimuli (vowel contrast differing in F_1, consonant contrast differing in VOT) being presented. As an alternative, for the Dutch /ɑ/–/aː/ vowel contrast, listeners rely heavily on both spectral and temporal information to categorize these sounds. Previous work from Reinisch and Sjerps (2013) has examined both SCEs and TCEs for this pair of vowels. Context sentences were manipulated both spectrally (high and low F_2) and temporally (fast and slow speaking rates), and listeners categorized minimal word pairs, which differed due to the /ɑ/–/aː/ contrast. Results indicated that identification of these vowels was influenced by both SCEs and TCEs. As such, testing the relationships between contextual variability, SCEs, and TCEs in the same target stimuli for Dutch-speaking listeners would provide the clearest test of these questions across acoustic domains.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional results from generalized linear mixed-effects model of responses in Experiment 1, Experiment 2, Experiment 3, across Experiments 1 and 2, and across Experiments 2 and 3.

ACKNOWLEDGMENTS

We thank Associate Editor Sven Mattys, Mark Pitt, and an anonymous reviewer for their valuable comments and feedback. We thank Isabel Adames, Pedro Gomez, Seth Long, Betsy Sellers, and Sydney Tharp for assistance with stimulus preparation and data collection. Experiment 1 comprised the second author's (C.M.S.) senior honors thesis in the Department of Psychological and Brain Sciences at the University of Louisville. Portions of this work were supported by National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant. No. R01 DC020303.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

This study was approved by the Institutional Review Board of the University of Louisville. All participants provided informed consent at the beginning of the experiment.

DATA AVAILABILITY

All data and annotated results scripts are available at https://osf.io/8wysv.

Assgari, A. A., and Stilp, C. E. (**2015**). "Talker information influences spectral contrast effects in speech categorization," J. Acoust. Soc. Am. **138**(5), 3023–3032.

Assgari, A. A., Theodore, R. M., and Stilp, C. E. (**2019**). "Variability in talkers' fundamental frequencies shapes context effects in speech perception," J. Acoust. Soc. Am. **145**(3), 1443–1454.

Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4. R package (version 1.1-33)," http://cran.r-project.org/package=lme4 (Last viewed February 23, 2024).

Boersma, P., and Weenink, D. (**2021**). "Praat: Doing phonetics by computer (version 6.2.17) [computer program]," http://www.praat.org (Last viewed February 23, 2024).

Bosker, H. R., and Ghitza, O. (**2018**). "Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation," Lang. Cogn. Neurosci. **33**(8), 955–967.

Bosker, H. R., Sjerps, M. J., and Reinisch, E. (**2020a**). "Spectral contrast effects are modulated by selective attention in 'cocktail party' settings," Atten. Percept. Psychophys. **82**, 1318–1332.

Bosker, H. R., Sjerps, M. J., and Reinisch, E. (**2020b**). "Temporal contrast effects in human speech perception are immune to selective attention," Sci. Rep. **10**(1), 5607.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**(3), 804–809.

Diehl, R. L., Souther, A. F., and Convis, C. L. (**1980**). "Conditions on rate normalization in speech perception," Percept. Psychophys. **27**, 435–443.

Dilley, L. C., and Pitt, M. A. (**2010**). "Altering context speech rate can cause words to appear or disappear," Psychol. Sci. **21**(11), 1664–1670.

Drown, L., and Theodore, R. M. (**2020**). "Effects of phonetic and indexical variability on talker normalization," J. Acoust. Soc. Am. **148**, 2504.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," National Institute of Standards and Technology, NIST Order No. PB91-505065.

Heffner, C. C., Dilley, L. C., McAuley, J. D., and Pitt, M. A. (**2013**). "When cues combine: How distal and proximal acoustic cues are integrated in word segmentation," Lang. Cogn. Processes **28**(9), 1275–1302.

Heffner, C. C., and Myers, E. B. (**2019**). "Variability in context effects on rate adaptation within individuals," J. Acoust. Soc. Am. **145**(3), 1790.

Holt, L. L. (**2005**). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci. **16**(4), 305–312.

Kahloon, L., Shorey, A. E., King, C. J., and Stilp, C. E. (**2023**). "Clear speech promotes speaking rate normalization," JASA Express Lett. **3**(5), 055205.

Kapadia, A. M., Tin, J. A. A., and Perrachione, T. K. (**2023**). "Multiple sources of acoustic variation affect speech processing efficiency," J. Acoust. Soc. Am. **153**(1), 209–223.

Kawahara, S., Kato, M., and Idemaru, K. (**2022**). "Speaking rate normalization across different talkers in the perception of Japanese stop and vowel length contrasts," JASA Express Lett. **2**(3), 035204.

Kidd, G. R. (**1989**). "Articulatory-rate context effects in phoneme identification," J. Exp. Psychol.: Hum. Percept. Perform. **15**(4), 736–748.

Kim, S., and Cho, T. (**2013**). "Prosodic boundary information modulates phonetic categorization," J. Acoust. Soc. Am. **134**(1), EL19–EL25.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104.

Laing, E. J. C., Liu, R., Lotto, A. J., and Holt, L. L. (**2012**). "Tuned with a tune: Talker normalization via general auditory processes," Front. Psychol. **3**(JUN), 203.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619.

Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol.: Hum. Percept. Perform. **33**(2), 391–409.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**(6), 457–465.

Mills, H. E., Shorey, A. E., Theodore, R. M., and Stilp, C. E. (**2022**). "Context effects in perception of vowels differentiated by F_1 are not influenced by variability in talkers' mean F_1 or F_3," J. Acoust. Soc. Am. **152**(1), 55–66.

Minifie, F. D., Kuhl, P. K., and Stecher, E. M. (**1977**). "Categorical perception of /b/ and /w/ during changes in rate of utterance," J. Acoust. Soc. Am. **62**(S1), S79.

Mitterer, H. (**2006**). "Is vowel normalization independent of lexical processing?," Phonetica **63**(4), 209–229.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**(5-6), 453–467.

Mullennix, J. W., Pisoni, D. B., and Martin, C. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**(1), 365–378.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**(4), 540–560.

Newman, R. S., and Sawusch, J. R. (**2009**). "Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another," J. Phon. **37**(1), 46–65.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099.

Oganian, Y., and Chang, E. F. (**2019**). "A speech envelope landmark for syllable encoding in human superior temporal gyrus," Sci. Adv. **5**(11) 1–13.

Oganian, Y., Kojima, K., Breska, A., Cai, C., Findlay, A., Chang, E., and Nagarajan, S. S. (**2023**). "Phase alignment of low-frequency neural activity to the amplitude envelope of speech reflects evoked responses to acoustic edges, not oscillatory entrainment," J. Neurosci. **43**(21), 3909–3921.

Pickett, J. M., and Decker, L. R. (**1960**). "Time factors in perception of a double consonant," Lang. Speech **3**(1), 11–17.

Reinisch, E. (**2016**). "Speaker-specific processing and local context information: The case of speaking rate," Appl. Psycholinguist. **37**(6), 1397–1415.

Reinisch, E., Jesse, A., and McQueen, J. M. (**2011**). "Speaking rate from proximal and distal contexts is used during word segmentation," J. Exp. Psychol.: Hum. Percept. Perform. **37**(3), 978–996.

Reinisch, E., and Sjerps, M. J. (**2013**). "The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context," J. Phon. **41**(2), 101–116.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. (**1978**). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol.: Hum. Percept. Perform. **4**(4), 621–637.

Samuel, A. G., and Pitt, M. A. (**2003**). "Lexical activation (and other factors) can mediate compensation for coarticulation," J. Mem. Lang. **48**(2), 416–434.

Sawusch, J. R., and Newman, R. S. (**2000**). "Perceptual normalization for speaking rate II: Effects of signal discontinuities," Percept. Psychophys. **62**(2), 285–300.

Shorey, A. E., and Stilp, C. E. (**2023**). "Short-term, not long-term, average spectra of preceding sentences bias consonant categorization," J. Acoust. Soc. Am. **153**(4), 2426–2435.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). "Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude," J. Acoust. Soc. Am. **96**(3), 1314–1324.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**(1), 112–117.

Stilp, C. E. (**2019**). "Auditory enhancement and spectral contrast effects in speech perception," J. Acoust. Soc. Am. **146**(2), 1503–1517.

Stilp, C. E. (**2020a**). "Acoustic context effects in speech perception," Wiley Interdiscip. Rev. Cogn. Sci. **11**(1), e1517.

Stilp, C. E. (**2020b**). "Evaluating peripheral versus central contributions to spectral context effects in speech perception," Hear. Res. **392**, 107983.

Stilp, C. E., and Assgari, A. A. (**2019**). "Natural signal statistics shift speech sound categorization," Atten. Percept. Psychophys. **81**(6), 2037–2052.

Stilp, C. E., and Assgari, A. A. (**2021**). "Contributions of natural signal statistics to spectral context effects in consonant categorization," Atten. Percept. Psychophys. **83**, 2694–2708.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol.: Hum. Percept. Perform. **7**(5), 1074–1095.

Szostak, C. M., and Pitt, M. A. (**2013**). "The prolonged influence of subsequent context on spoken word recognition," Atten. Percept. Psychophys. **75**, 1533–1546.

Watkins, A. J. (**1991**). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**(6), 2942–2955.

Winn, M. B. (**2020**). "Accommodation of gender-related phonetic differences by listeners with cochlear implants and in a variety of vocoder simulations," J. Acoust. Soc. Am. **147**(1), 174–190.

Winn, M. B., Won, J. H., and Moon, I. J. (**2016**). "Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization," Ear. Hear. **37**(6), e377–e390.