1 Experiment 2: Stop-glide perception

This experiment replicated Miller & Liberman's (1979) study of the effects of following vowel duration on the perception of transition duration as a cue to the /b:w/ contrast, with two modifications. First, varying the duration of the following vowel altered whether it was perceived as one nucleus or two. This manipulation tested the hypothesis that listeners would perceive the following vowel as shorter when they perceived it as the first of two syllable nuclei. If so, they were predicted to respond "w" more often when the following vowel was long, not less often. The opposite of Miller & Liberman's finding was predicted because a longer vowel was more likely to be parsed into two nuclei, and only the perceived duration of the first would serve as the context for interpreting the transition duration. Because the perceived duration of that first nucleus was only a fraction of the total duration of the vowel, a shorter transition duration was predicted to perceived as long enough for listeners to respond "w," the long transition duration category. Second, we again manipulated the rate of the sentence containing the test items. Experiment 1 showed that as the rate of speaking increased, listeners were more likely to hear the vowel as two nuclei. The vowels that were judged to consist of two nuclei were physically shorter in the context of a sentence spoken at a faster rate than in one spoken at a slower rate (Figure reffig: observedExp01), so this manipulation tests whether listeners respond "w" at yet shorter vowel durations at faster rates because a shorter vowel duration was more likely to be parsed as two nuclei. Refer to Summerfield 81?

1.1 Methods

1.1.1 Stimuli

All the stimuli were constructed from additional sentences recorded by the same adult male native English speaker who produced the sentences for Experiment 1. They were recorded at a sampling rate of 44.1 kilohertz with 16 bit resolution. The sentences were "The butler was killed by Mr. Byatt in the library," and "The butler was killed by Mr. Wyatt in the library." Progressively longer intervals from the initial 90 ms interval from "Byatt" were replaced with corresponding intervals from "Wyatt," starting with the stop burst, until a 10 step continuum from /b/ to /w/ was created. The portions of the initial interval from "Byatt" replaced with corresponding portions from "Wyatt" lasted 5, 14, 27, 36, 45, 55, 64, 74, 81, and 90 ms. All junctures between the portions from "Wyatt" and the remainder of "Byatt" were at zero crossings between successive periods.

These ten intervals were then concatenated with four of the ensuing vowel durations used in the Experiment 1: the shortest (unambiguously one nucleus), the longest (unambiguously two nuclei), and the two steps that straddled the category boundary between one and two nuclei at the medium rate of speaking (steps 70% and 80% in the 11 step vowel continuum, see Figure **??**). Figure 1 displays the most /b/-like and the most /w/-like initial intervals together with an intermediate step along this continuum (step 5) concatenated with an intermediate duration vowel continuation (step 70%). The resulting names were then inserted into the same fast, medium, and slow sentence frames used in Experiment 1. Other characteristics of the stimuli replicated those used in Experiment 1.

1.1.2 Listeners

Twelve native speakers of American English were recruited by advertisement and received course credit or payment ($10/hour) for their participation. They reported no speaking or listening disorder. None of them had participated in Experiment 1.

1.1.3 Procedures

Experiment 2 began with unscored endpoint training with feedback. The training consisted of the two most /b/-like transition durations and the two most /w/ transition durations attached to the longest and shortest vowels in all three sentence/rate contexts. Each combination was presented twice, for a total of 48 training trials (4 endpoints × 2 vowel durations × 3 contexts × 2 repetitions). After the listener responded, or following 1500 ms after the completion of the sentence, the correct response was displayed on the screen for 750 ms. This feedback was given during the training block only. The next stimulus was presented 750 ms later.

Training was followed by eight test blocks in total. Half of the blocks consisted of the odd-numbered stimuli from the transition duration continuum, the other half of the even-numbered stimuli. In each block, the stimuli at or next to one of the endpoints (1 or 2 and 9 or 10) were presented once, and the three intermediate stimuli (3, 5, 7 or 4, 6, 8) were presented three times. Thus, each stimulus near the endpoints was presented a total of 4 times, and each intermediate stimulus 12 times. As each transition duration was combined with four following vowel durations and these four combinations are presented in sentences with three speaking rates, a block consisted of 132 trials (11 samples from the transition duration continuum × 4 vowel durations × 3 speaking rates).

After each test block, which lasted approximately 9 minutes, depending on response times, a message was displayed on the screen that instructed listeners to take a short break and to push a button when they were ready to continue on to the next block. Halfway through the experiment, a message appeared on the screen which instructed the listener to to get up and leave the listening room. This mandatory break lasted about 5 minutes. The experiment lasted approximately 90 minutes altogether, including briefing beforehand and debriefing afterward.

Each test trial presented the listener with a sentence and four options for labeling the target name: “Bight,” “Byatt,” “Wight,” or “Wyatt,” arranged as a 2 x 2 array on the screen. The color and orientation of the visual prompt words corresponded to the arrangement and color of the buttons on the response pad. During the entire playing of the sentence, the four choices were left on the screen, and the listener was instructed to respond as soon as they were sure of their choice. As soon as they responded, the sentence stopped playing and the 750 ms inter-trial interval began. If listeners did not respond before the sentence ended, they were given up to 1500 ms afterward to respond. If after 1500 ms passed they had not responded, the inter-trial interval began, after which the next stimulus began.

The experiment had four conditions, which were designed to compensate for the location of the buttons corresponding to different responses. The two “b” names were either assigned to the two left, the two right, the two top, or the two bottom buttons, and the two “w” names to the other pair. Three participants was run on each of the four conditions, for a total of 12.

1.2 Results

The results are presented here in a form that matches the design, but that might appear at first glance unconventional. There were four possible responses on each trial, “Bight,” “Byatt,” ”Wight,” and “Wyatt.” As they were presented to listeners they form a 2 by 2 matrix, with dimensions “one” versus “two” syllables and “b” versus “w.” This response structure is neither multinomial – the responses are not independent categories but instead categories cross-classified by the number of syllables and the manner of the consonant – nor is it ordi-

nal – "bight" might appear to be the lowest rank and "wyatt" the highest, but any ranking of "byatt" and "wight" would be arbitrary or worse *ad hoc*. The proportions of the four possible responses were therefore cross-classified for both categories at once, such that the long responses, two syllables or glide, were coded as 1 and the short ones, one syllable or stop as 0. These response proportions were submitted to a hierarchy of multilevel, multivariate logistic regression models using the *brms* package (Bürkner, 2017,?) with uninformative priors. Population-level effects were the scaled duration of the transition, the scaled duration of the vowel, and the scaled rate of the sentence (slow = -1, medium = 0, fast = 1). Decorrelated group-level effects of listener were included on the intercept and the slopes of transition duration, vowel duration, and rate, but not on any interactions.

Figures 3 and 3 show the proportions of the four possible responses by VOT and sentential rate or vowel duration. Both show that "bight" and "byatt" responses decreased and "wight" and "wyatt" responses increased as transition duration values increased. Figure 3 shows that one-syllable "bight" and "wight" responses decreased and two-syllable "byatt" and "wyatt" responses increased as sentential rate sped up. Figure 3 shows that one-syllable "bight" and "wight" responses decreased and two-syllable "byatt" and "wyatt" responses increased as the following vowel lengthened.

Two late models in this hierarchy are compared here, one without any interactions between the population-level effects, and one with all two- and three-way interactions. Leave-one-out comparison of LOOIC values showed that the more complex model including these interactions did improve on the simpler one lacking them, LOOIC = 16825.0 (se = 213.2) versus 16871.8 (se = 212.7). The large standard errors of these estimates exceed the difference between them show, however, that improvement is modest. Nonetheless, the estimates of population-level effects from the more complex model are reported here because the 95% credible intervals of a number of interactions did not include 0.

The population-level estimates in Table 1 show that listeners favored "w" over "b" responses (positive w:b intercept) and "one syllable" over "two syllables" responses (negative 2:1 intercept). They also responded "w" more often as the transition duration increased (positive w:b transition), but less often as vowel duration increased and in a sentence with a faster rate (negative w:b vowel and rate). A faster rate also reduced the "w" responses for longer transitions (negative w:b transition by rate interaction), but increased them for longer transitions at longer vowel durations (positive w:b transition by vowel by rate interaction). Listeners responded "two syllables" more often when the vowel was longer and the rate faster (positive 2:1 vowel and rate), but less often when the transition was longer (negative 2:1 transition). Yet "two syllable" responses increased more for longer vowels when the transition was also longer (positive 2:1 transition by vowel interaction). The variation between listeners reflected in the group-level standard deviations was the same order of magnitude as the corresponding population-level effects. All of the 95% credible intervals for correlations in the random effects between the responses included 0, which shows that variation between listeners' biases and sensitivity to stimulus properties for one response was unrelated to those for the other response.

1.3 Discussion

Listeners' responses conformed to expectations regarding the effects of the principal cues: "w" responses increased with transition durations and "two-syllable" responses increased with vowel duration. As expected from Miller & Liberman's (1979) results, "w" responses instead decreased when the following vowel was longer. The decrease in the effect of transition du-

Table 1: Estimates, estimated errors, and the lower and upper bounds of the 95% credible intervals of the population-level effects from the model including the interaction between transition duration, vowel duration, and rate, and including group-level effects on the intercept and the slopes of duration and rate.

Effect	Population-level				Group-level standard deviations			
	Estimate	Estimated Error	lower bound	upper bound	Estimate	Estimated Error	lower bound	upper bound
w:b intercept	1.17	0.37	0.42	1.90	1.28	0.30	0.83	2.00
w:b transition	4.24	0.83	2.59	5.91	2.78	0.60	1.85	4.18
w:b vowel	-0.15	0.05	-0.26	-0.05	0.13	0.07	0.01	0.27
w:b rate	-0.15	0.09	-0.34	0.04	0.25	0.08	0.13	0.44
w:b trans:vowel	-0.10	0.05	-0.20	-0.00				
w:b trans:rate	-0.22	0.05	-0.32	-0.12				
w:b vowel:rate	-0.01	0.03	-0.07	0.06				
w:b trans:vowel:rate	0.12	0.05	0.03	0.22				
2:1 intercept	-0.29	0.10	-0.48	-0.09	0.30	0.08	0.18	0.49
2:1 transition	-0.53	0.12	-0.77	-0.28	0.40	0.11	0.25	0.66
2:1 vowel	2.73	0.37	2.00	3.47	1.24	0.27	0.82	1.88
2:1 rate	0.84	0.10	0.66	1.03	0.31	0.08	0.20	0.49
2:1 trans:vowel	0.22	0.04	0.14	0.30				
2:1 trans:rate	0.05	0.03	-0.00	0.11				
2:1 vowel:rate	-0.06	0.04	-0.14	0.02				
2:1 trans:vowel:rate	-0.01	0.04	-0.09	0.06				

rations at faster rates is, however, unexpected if listeners normalize for rate, as a shorter transition duration should be long enough to convey "w" when speech is perceived to be fast. The increase in "two-syllable" responses at faster rates replicated that found in Experiment 1. The decrease in "two-syllable" responses as transition duration increased could be a product of listeners' attributing part of the vowel to glide *** is that compatible with the decrease in "w" responses as the vowel duration increased?

References

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models. *Journal of Statistical Software*, *80*(1), 1-28.

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

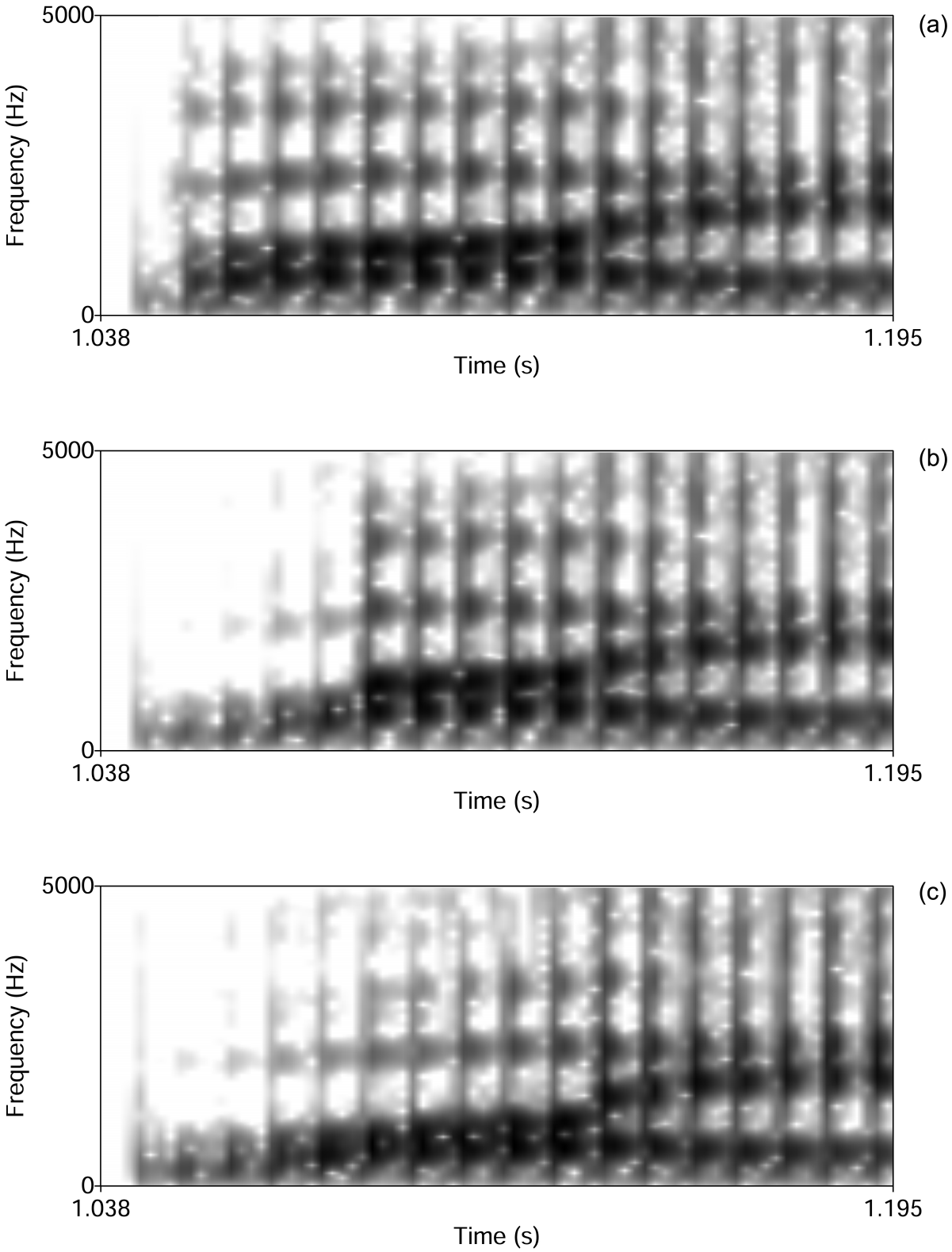

Figure 1: Steps 1, 5, and 10 of the 10-step /b-w/ transition duration continuum, combined with step 70% from the vowel duration continuum from Experiment 1.

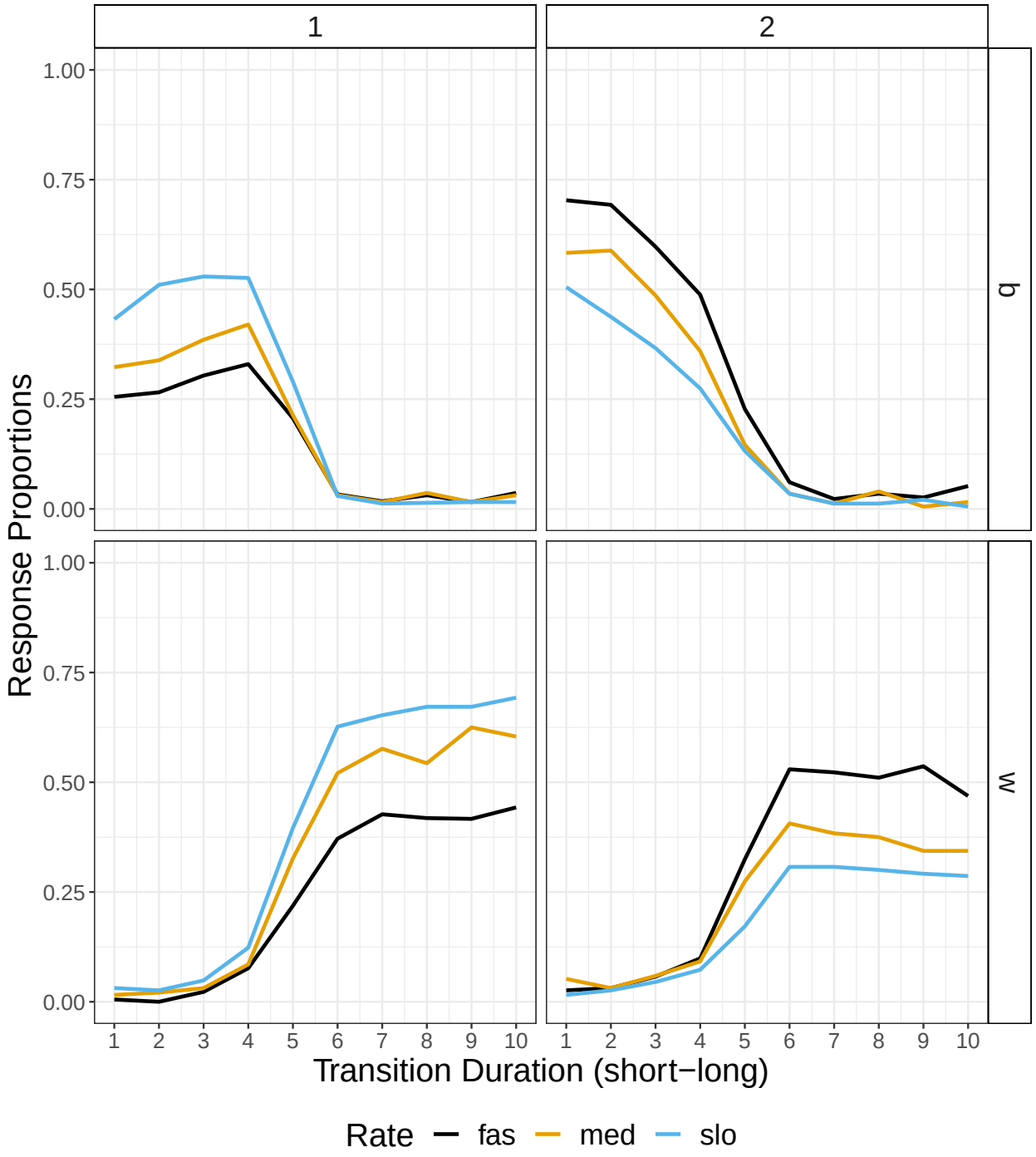

Figure 2: Mean proportion of "bight" (upper left), "byatt" (upper right), "wight" (lower left), and "wyatt" (lower right) responses by transition duration, as a function of sentential rate (colors).

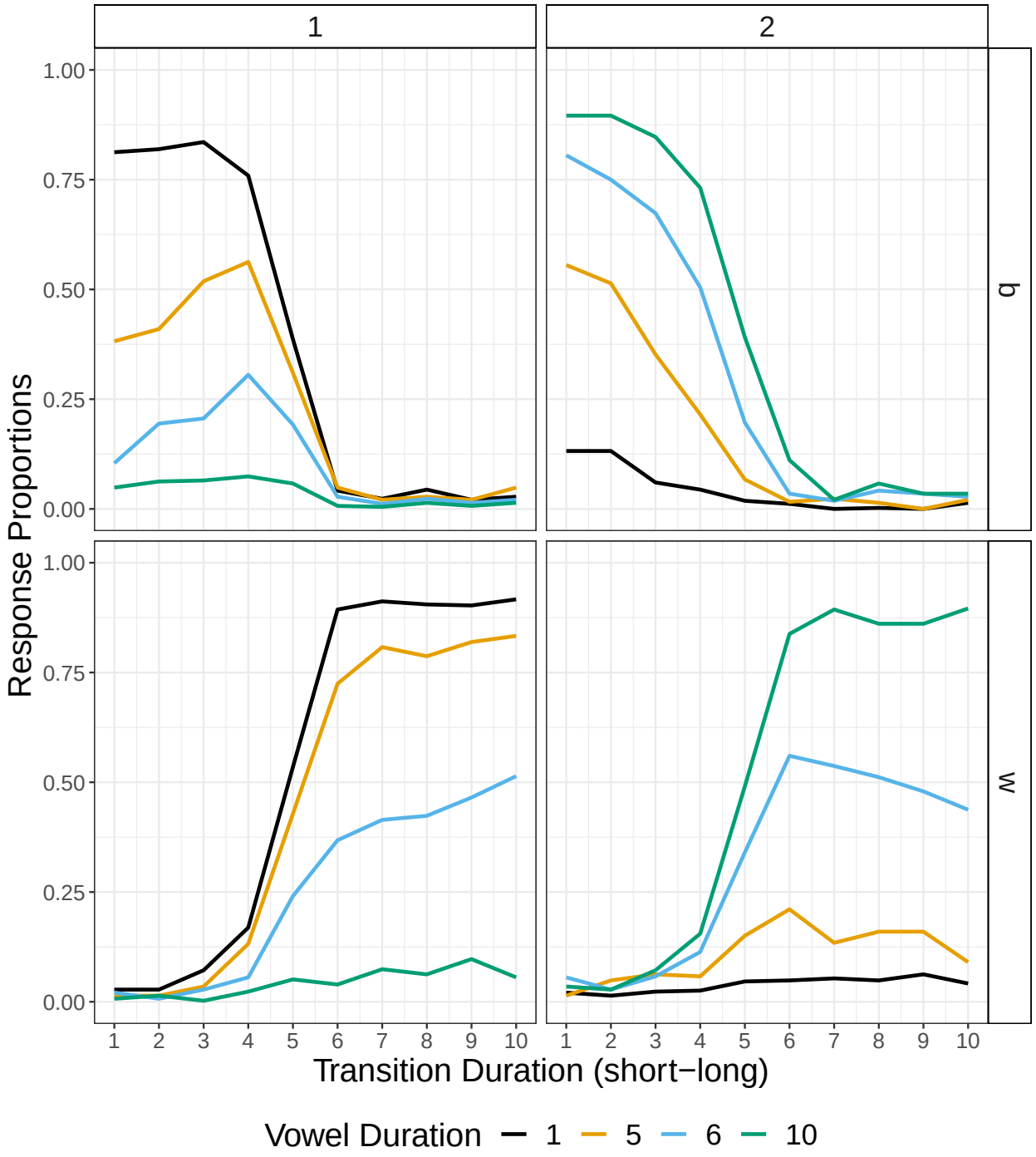

Figure 3: Mean proportion of "bight" (upper left), "byatt" (upper right), "wight" (lower left), and "wyatt" (lower right) responses by transition duration, as a function of vowel duration (colors).