Appendix

1.0 Acoustic measurements for each stimulus item

The tables within the body of the paper provide summary acoustics for stimuli divided by the original coda voicing environment. The tables here further separate the stimuli, to provide the characteristics separated by coda voicing and the other manipulations, with the exception of duration step. The ten duration steps are pooled together within each category. Note that the manipulations can interact with each other, so even with the same starting recording, the outputs may differ.

Some differences across stimuli reflect systematic effects, but other variation merely reflects idiosyncratic characteristics of particular recordings. These measurements are thus only meant to provide further information about the particular stimuli used in the experiments; they are not meant to provide production data to demonstrate interacting influences on acoustic characteristics of vowels. Refer to the body of the paper for references to relevant production studies.

Table 1 presents a summary of the acoustic characteristics of the stimuli used in Experiment 1, separated by each of the main factors other than duration step.

Table 1. Acoustic characteristics of vowels in the Experiment 1 stimuli, by vowel quality, F1 manipulation and the original coda's voicing

			f0 mean	f0 Δ (Q4-Q1)	intensity Δ (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	æ	low F1	196 Hz	2.5 Hz	1.2 dB	788 Hz	-20.0 dB	0.6%	15.7
Voiced	æ	mid F1	196 Hz	2.7 Hz	0.7 dB	883 Hz	-19.0 dB	0.7%	13.7
Voiced	æ	high F1	196 Hz	2.6 Hz	1.1 dB	969 Hz	-19.0 dB	0.6%	13.7
Voiced	ɛ	low F1	224 Hz	3.1 Hz	-1.0 dB	727 Hz	-6.3 dB	1.0%	13.7
Voiced	ɛ	mid F1	224 Hz	2.9 Hz	-1.1 dB	842 Hz	-6.1 dB	0.6%	14.8
Voiced	ɛ	high F1	224 Hz	2.8 Hz	-0.8 dB	872 Hz	-6.0 dB	0.7%	13.7
Voiced	ɪ	low F1	218 Hz	8.0 Hz	3.6 dB	488 Hz	-13.7 dB	0.4%	15.7
Voiced	ɪ	mid F1	218 Hz	8.4 Hz	3.8 dB	535 Hz	-12.3 dB	0.4%	14.3
Voiced	ɪ	high F1	218 Hz	8.3 Hz	3.7 dB	693 Hz	-5.9 dB	0.4%	13.4
Voiceless	æ	low F1	210 Hz	-3.0 Hz	-0.1 dB	847 Hz	-11.4 dB	1.0%	16.0
Voiceless	æ	mid F1	210 Hz	-2.4 Hz	-0.4 dB	933 Hz	-10.3 dB	0.7%	14.5
Voiceless	æ	high F1	210 Hz	-2.4 Hz	-0.5 dB	1137 Hz	-9.8 dB	0.8%	12.3
Voiceless	ɛ	low F1	231 Hz	14.1 Hz	-0.4 dB	819 Hz	-4.1 dB	0.3%	15.8
Voiceless	ɛ	mid F1	231 Hz	13.8 Hz	-0.3 dB	873 Hz	-4.2 dB	0.3%	15.4
Voiceless	ɛ	high F1	231 Hz	13.7 Hz	-0.4 dB	900 Hz	-4.3 dB	0.4%	15.5
Voiceless	ɪ	low F1	256 Hz	13.6 Hz	-0.7 dB	531 Hz	-15.5 dB	0.4%	16.6
Voiceless	ɪ	mid F1	256 Hz	14.4 Hz	-1.4 dB	631 Hz	-12.3 dB	0.7%	13.0
Voiceless	ɪ	high F1	256 Hz	14.6 Hz	-1.3 dB	690 Hz	-11.4 dB	0.8%	12.8

Table 2 presents a summary of the acoustic characteristics of the stimuli used in Experiment 2, separated by each of the main factors other than duration step.

Table 2. Acoustic characteristics of vowels in the Experiment 2 stimuli, by intensity manipulation and the original coda's voicing

		f0 mean	f0 Δ (Q4-Q1)	intensity Δ (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	falling intensity	192 Hz	-0.1 Hz	-6.1 dB	860 Hz	6.8 dB	1.8%	9.4
Voiced	level intensity	195 Hz	-9.9 Hz	1.8 dB	871 Hz	1.2 dB	1.5%	6.8
Voiced	rising intensity	195 Hz	-12.7 Hz	10.3 dB	918 Hz	0.7 dB	0.6%	9.8
Voiceless	falling intensity	199 Hz	-8.1 Hz	-16.2 dB	1047 Hz	6.2 dB	1.1%	8.2
Voiceless	level intensity	198 Hz	-6.8 Hz	0.3 dB	1083 Hz	5.9 dB	1.1%	9.1
Voiceless	rising intensity	198 Hz	-8.7 Hz	11.3 dB	1040 Hz	6.2 dB	2.2%	7.1

Table 3 presents a summary of the acoustic characteristics of the stimuli used in Experiment 3, separated by each of the main factors other than duration step.

Table 3. Acoustic characteristics of vowels in the Experiment 3 stimuli, by spectral tilt manipulation and the original coda's voicing

		f0 mean	f0 Δ (Q4-Q1)	intensity Δ (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	low spectral tilt	168 Hz	-18.4 Hz	-0.3 dB	767 Hz	-1.1 dB	1.0%	5.8
Voiced	mid spectral tilt	168 Hz	-20.0 Hz	1.5 dB	680 Hz	3.0 dB	1.0%	9.3
Voiced	high spectral tilt	168 Hz	-20.0 Hz	1.0 dB	642 Hz	6.4 dB	1.0%	11.9
Voiceless	low spectral tilt	167 Hz	-23.7 Hz	1.1 dB	869 Hz	1.2 dB	1.2%	2.9
Voiceless	mid spectral tilt	167 Hz	-24.9 Hz	-0.04 dB	776 Hz	5.0 dB	1.6%	6.7
Voiceless	high spectral tilt	168 Hz	-25.0 Hz	-2.3 dB	679 Hz	8.9 dB	1.4%	10.2

Table 4 presents a summary of the acoustic characteristics of the stimuli used in Experiment 4, separated by each of the main factors other than duration step.

Table 4. Acoustic characteristics of vowels in the Experiment 4 stimuli, by vowel quality and the original coda's voicing

	vowel	f0	f0 Δ (Q4-Q1)	intensity Δ (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	a	206 Hz	50.9 Hz	-0.6 dB	914 Hz	-4.7 dB	1.1%	7.2
Voiced	i	233 Hz	53.1 Hz	9.8 dB	382 Hz	13.9 dB	1.3%	10.9
Voiced	u	246 Hz	23.4 Hz	3.7 dB	427 Hz	2.4 dB	1.1%	15.8
Voiceless	a	200 Hz	22.1 Hz	0.2 dB	889 Hz	-1.9 dB	0.7%	10.1
Voiceless	i	245 Hz	18.5 Hz	3.9 dB	477 Hz	10.9 dB	0.5%	11.4
Voiceless	u	244 Hz	65.2 Hz	5.3 dB	461 Hz	4.0 dB	1.6%	12.1

2.0 Interactions with duration step

Some of the effects interact with duration step. This section presents models that include interactions between duration step and the main manipulations of interest. Other interactions are not included, both because models including every interaction are unlikely to converge, and moreover because testing a large number of models requires adjustments for significance. These interactions were not addressed in the original hypotheses, so any post-hoc analyses should be examined cautiously. It is not clear why the interactions are significant in some experiments but not in others; additional work would be necessary to properly address this question.

As can be seen from the figures in the main body of the paper, most of the interactions with duration seem to indicate that influences on perceived duration are weaker at the longest durations, the shortest durations, or both. This might be predicted simply from ceiling and floor effects; when there is little variability in how a duration is perceived, it will be harder to see evidence for factors influencing identifications. Such an explanation also predicts that the interaction does not have a linear effect, so these models might not entirely capture it.

Table 5 presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 1. The fixed effects were vowel duration step; vowel quality (/æ, ɛ, ɪ/); F1 step within the vowel; voicing of the original coda (voiced, voiceless); and the interaction between duration step and vowel quality. There was a random intercept for participant, and there were no random slopes. Duration step was centered, so that the βs for the interaction with vowel quality would be interpretable values.

Table 5. Regression model for 'long' responses, Experiment 1. *Reference Levels: Vowel = ɛ; OrigCoda = Voiced*

	β	*SE*	z value	p value
(Intercept)	-0.195	0.194.	-1.01	0.314
Duration Step	0.572	0.029	19.7	< 0.001***
Vowel /ɪ/	0.375	0.0948	3.96	< 0.001***
Vowel /æ/	-0.0326	0.0951	-0.343	0.732
F1	0.0133	0.0467	0.285	0.776
OrigCoda Voiceless	-0.17	0.0763	-2.23	0.0257*
Vowel /ɪ/ * Step	-0.0974	0.0383	-2.55	0.0109*
Vowel /æ/ * Step	-0.0976	0.0384	-2.55	0.0109*

The interaction between vowel quality and duration step does have an effect: The difference between the probability of 'long' responses for /ɪ/ and /ɛ/ is smaller at longer durations. The number of long responses for /æ/, relative to /ɛ/, also decreases at longer durations.

The results for the main effects are the same as they were in the original model: 'Long' responses increase with longer durations, 'long' responses are more frequent for /ɪ/ than /ɛ/, 'long' responses are less frequent with stimuli made from recordings with voiceless codas, and there is no effect of F1.

Table 6 presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 2. The fixed effects were vowel duration step; intensity contour (rising, falling, level); voicing of the original coda (voiced, voiceless); the voicing of the spliced ending (voiced, voiceless); and the interaction between duration step and intensity. There was a random intercept for participant, and there were no random slopes. Duration step

was centered, so that the βs for the interaction with intensity would be interpretable values.

Table 6. Regression model for 'long' responses, Experiment 2. *Reference Levels: Intensity = Level; OrigCoda = Voiced; Ending = Voiced*

	β	*SE*	z value	p value
(Intercept)	-0.415	0.124	-3.36	< 0.001***
Duration Step	0.379	0.029	13.1	< 0.001***
Intensity Rising	0.397	0.107	3.72	< 0.001***
Intensity Falling	-0.51	0.107	-4.78	< 0.001***
OrigCoda Voiceless	-0.471	0.0867	-5.44	< 0.001***
Ending Voiceless	1.24	0.0889	14.0	< 0.001***
Intensity Rising * Step	-0.0705	0.0395	-1.78	0.0744
Intensity Falling * Step	-0.12	0.0392	-3.06	0.00219**

The interaction between intensity contour and duration step has an effect: The difference between the probability of 'long' responses for rising and level intensity items is smaller at longer durations. The number of 'long' responses for falling intensity items, relative to level intensity items, also decreases with longer durations.

The results for the main effects are the same as they were in the original model: 'Long' responses increase with longer durations, 'long' responses are more frequent with rising intensity than level intensity and less frequent with falling intensity, 'long' responses are less frequent with stimuli made from recordings with voiceless codas, and 'long' responses are more frequent when there is a voiceless ending spliced onto the vowel.

Table 7 presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 3. The fixed effects were vowel duration step; spectral tilt (high, low, moderate); voicing of the original coda (voiced, voiceless); voicing of the spliced ending (voiced, voiceless); and the interaction between duration step and spectral tilt. There was a random intercept for participant, and there were no random slopes. Duration step was centered, so that the βs for the interaction with spectral tilt would be interpretable values.

Table 7. Regression model for 'long' responses, Experiment 3. *Reference Levels: Spectral tilt = Moderate; OrigCoda = Voiced; Ending = Voiced*

	β	*SE*	z value	p value
(Intercept)	0.259	0.147	1.76	0.0778
Duration Step	0.416	0.0213	19.5	< 0.001***
Spectral Tilt High	-0.333	0.0763	-4.36	< 0.001***
Spectral Tilt Low	0.141	0.0781	1.8	0.0717
OrigCoda Voiceless	-0.209	0.0626	-3.35	< 0.001***
Ending Voiceless	0.199	0.0626	3.18	0.00148**
Tilt High * Step	-0.04	0.0292	-1.37	0.17
Tilt Low * Step	0.0212	0.0304	0.697	0.486

The interaction between spectral tilt and duration step does not significantly improve the model ($\chi^2 = 4.42$, $df = 2$, $p = 0.11$).

The results for the main effects are the same as they were in the original model: 'Long' responses increase with longer durations, 'long' responses are less frequent with high spectral tilt than moderate spectral tilt, 'long' responses are less frequent with stimuli made from recordings with voiceless codas, and 'long' responses are more frequent when there is a voiceless ending spliced onto the vowel.

Table 8 presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 4. The fixed effects were vowel duration step; vowel quality (/a, i, u/); voicing of the original coda (voiced, voiceless); and the interaction between duration step and voicing of the original coda. There was a random intercept for participant, and there were no random slopes. Duration step was centered, so that the βs for the interaction with voicing of the original coda would be interpretable values.

Table 8. Regression model for 'long' responses, Experiment 4. *Reference Levels: Vowel = a; OrigCoda = Voiced*

	β	*SE*	*z* value	p value
(Intercept)	0.0141	0.137	0.102	0.919
Duration Step	0.516	0.0274	18.8	< 0.001***
Vowel /i/	0.119	0.112	1.06	0.288
Vowel /u/	0.175	0.112	1.56	0.118
OrigCoda Voiceless	-0.245	0.0918	-2.67	0.00757
OrigCoda Voiceless * Step	-0.041	0.0373	-1.1	0.272

The interaction between the voicing of the original coda and duration step does not significantly improve the model ($\chi^2 = 1.21$, $df = 1$, $p = 0.272$).

The results for the main effects are the same as they were in the original model: 'Long' responses increase with longer durations, 'long' responses are less frequent with stimuli made from recordings with voiceless codas, and vowel quality doesn't have a significant effect.

Including the interactions does not change the main effects in any model. Thus, the inclusion or exclusion of these interactions does not change the results reported in the paper.

3.0 Specific spliced codas in Experiment 3

The model for Experiment 3 presented in the body of the paper only includes voicing of the spliced coda as a factor, rather than dividing the spliced codas into the four endings used in the experiment (/b, d, p, t/). Separating this factor into the four specific endings did not produce a model with a better fit than one which instead grouped the endings based on voicing ($\chi^2 = 0.431$, $df = 2$, $p = 0.806$). That model is nonetheless presented here in Table 9. This table presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 3. The fixed effects were vowel duration step; spectral tilt (high, low, moderate); voicing of the original coda (voiced, voiceless); and the spliced ending (/b, d, p, t/). There was a random intercept for participant, and there were no random slopes.

As compared to the reference level /t/, the model demonstrates that both a spliced coda /b/ and coda /d/ elicit fewer long responses (the same effect captured by a factor that only includes voicing). In contrast, coda /p/ does not differ from /t/.

Table 10 provides an alternative model in which voicing and place of articulation are included as separate factors. This table presents the summary of a mixed effects logistic regression

Table 9. Regression model for 'long' responses, Experiment 3. *Reference Levels: Spectral tilt = Moderate; OrigCoda = Voiced; Ending = t*

	β	*SE*	z value	p value
(Intercept)	-1.8	0.166	-10.9	< 0.001***
Duration Step	0.409	0.0125	32.6	< 0.001***
Spectral Tilt High	-0.333	0.0766	-4.35	< 0.001***
Spectral Tilt Low	0.132	0.0768	1.72	0.0858
OrigCoda Voiceless	-0.21	0.0626	-3.36	< 0.001***
Ending b	-0.163	0.0883	-1.85	0.065
Ending d	-0.218	0.0882	-2.47	0.0134*
Ending p	0.018	0.0886	0.203	0.839

model for the 'long' responses to each item in Experiment 3. The fixed effects were vowel duration step; spectral tilt (high, low, moderate); voicing of the original coda (voiced, voiceless); the voicing of the spliced ending (voiced, voiceless); and the place of articulation of the spliced ending (alveolar, labial). There was a random intercept for participant, and there were no random slopes.

Table 10. Regression model for 'long' responses, Experiment 3. *Reference Levels: Spectral tilt = Moderate; OrigCoda = Voiced; EndingVoice = Voiced; EndingPlace = Alveolar*

	β	*SE*	z value	p value
(Intercept)	-2.01	0.164	-12.3	< 0.001***
Duration Step	0.409	0.0125	32.6	< 0.001***
Spectral Tilt High	-0.333	0.0766	-4.35	< 0.001***
Spectral Tilt Low	0.132	0.0768	1.72	0.0855
OrigCoda Voiceless	-0.21	0.0626	-3.36	< 0.001***
EndingVoice Voiceless	0.2	0.0626	3.19	0.00142**
EndingPlace Labial	0.0366	0.0625	0.586	0.558

This model similarly demonstrates that there is an effect of the voicing of the spliced coda on the number of 'long' responses, but no effect of the place of articulation of the spliced coda.

The reason why none of the models find an effect of place of articulation can also be seen in the mean proportion of 'long' responses in each category, presented in Table 11: There is almost no difference between /t/ and /p/, and very little difference between /d/ and /b/.

Table 11. Proportion of 'long' responses for stimuli with each spliced coda

p	t	b	d
0.553	0.553	0.521	0.511

4.0 Additional figures

The main body of the paper only provides figures for the main factor or factors of interest for each study, in consideration of space. Figures are provided in this section for the factors included in the regression models but not visualized within the paper. Some of the effects are small, and look even smaller when separated by duration step, as they are here. The effects and their sizes can be identified more clearly in the outputs of the regression models in the main body of the paper.

Figure 1 illustrates the effects of voicing of the original coda on perceived duration in Experiment 1.

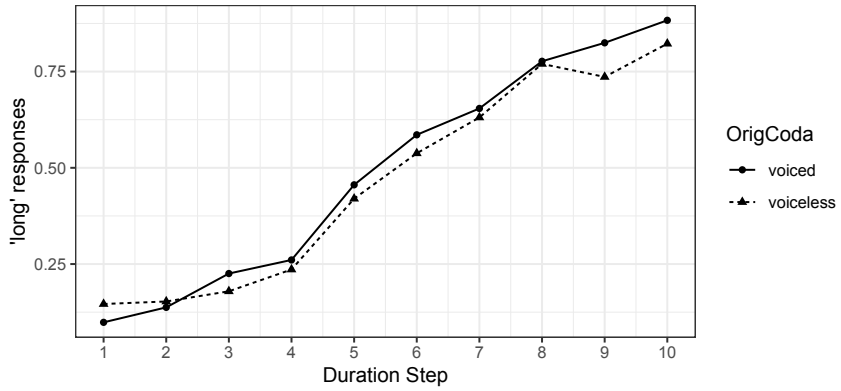


Figure 1. Proportion of 'long' responses in Experiment 1, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model, and pooled across participants.

Figure 2 illustrates the effects of voicing of the original coda on perceived duration in Experiment 2.

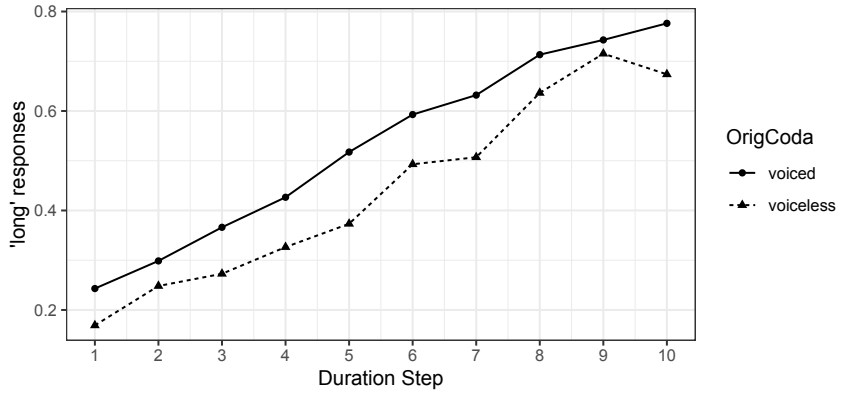


Figure 2. Proportion of 'long' responses in Experiment 2, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model, and pooled across participants.

Figure 3 illustrates the effects of voicing of the spliced ending on perceived duration in Experiment 2.

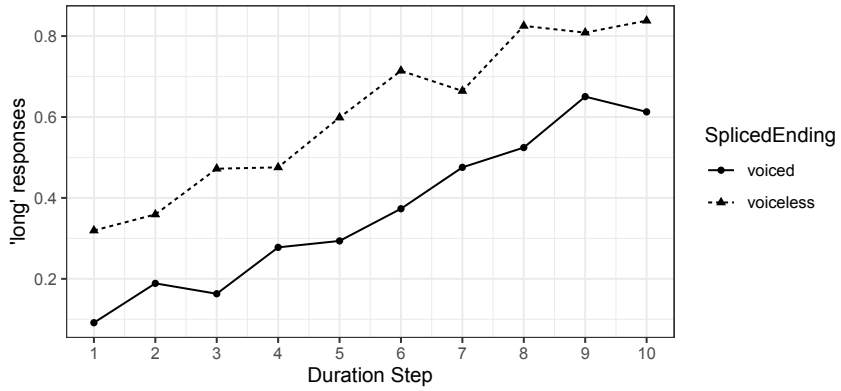


Figure 3. Proportion of 'long' responses in Experiment 2, by duration step and voicing of the spliced ending. Based on the raw data, not the output of the regression model, and pooled across participants.

Figure 4 illustrates the effects of voicing of the original coda on perceived duration in Experiment 3.

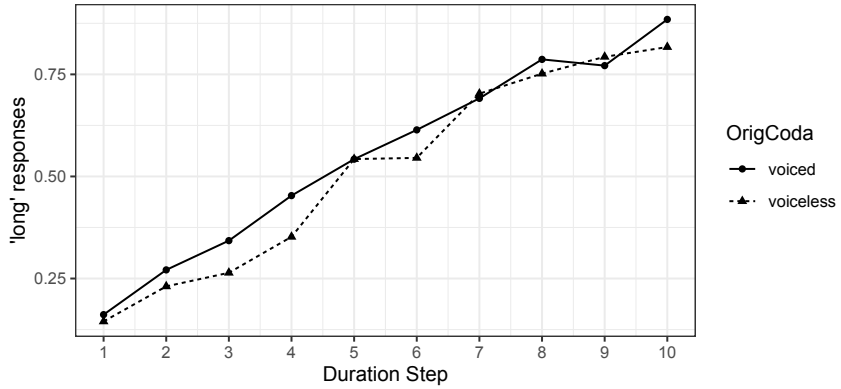


Figure 4. Proportion of 'long' responses in Experiment 3, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model, and pooled across participants.

Figure 5 illustrates the effects of voicing of the spliced ending on perceived duration in Experiment 3.

Figure 6 illustrates the effects of vowel quality on perceived duration in Experiment 4.

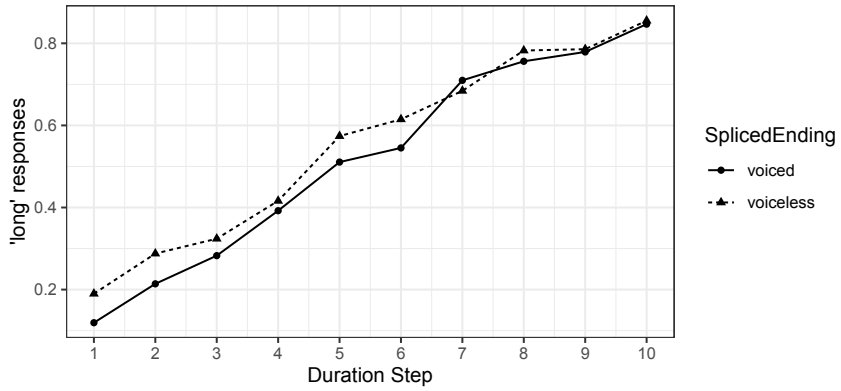


Figure 5. Proportion of 'long' responses in Experiment 3, by duration step and voicing of the spliced ending. Based on the raw data, not the output of the regression model, and pooled across participants.

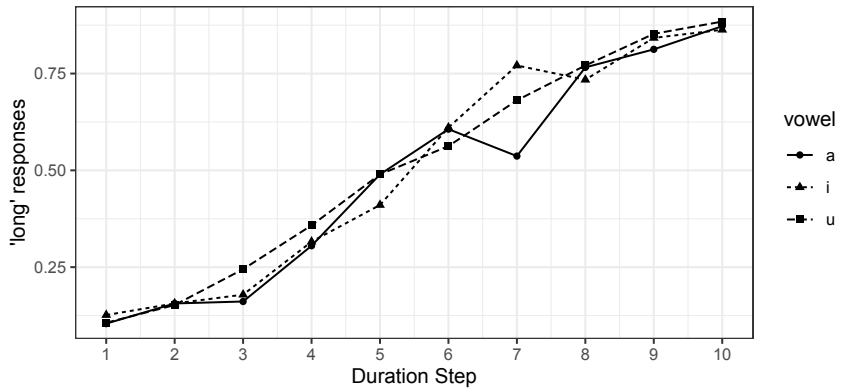
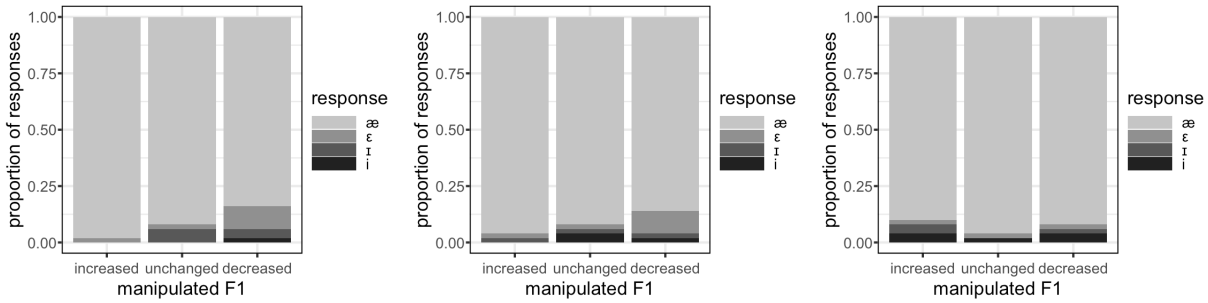


Figure 6. Proportion of 'long' responses in Experiment 4, by duration step and vowel quality. Based on the raw data, not the output of the regression model, and pooled across participants.

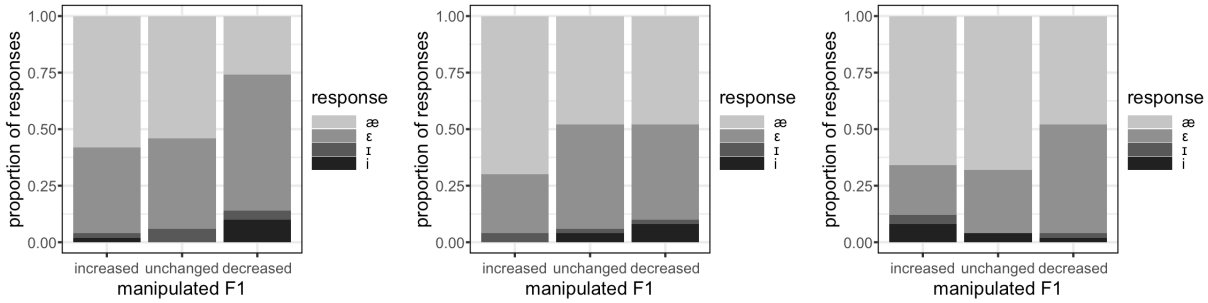
5.0 Results for the follow-up to Experiment 1, separated by F1 and duration

Within the main body of the paper, the results for the follow-up to Experiment 1 are presented separately for F1 and for duration, in each case pooling by the characteristic not being analyzed. Figures 7-9 present both factors together. Each column of results in each of the figures is presenting responses for stimuli with a particular vowel quality, F1 manipulation, and duration manipulation; thus, each column represents just two stimulus items, from the two original coda environments.



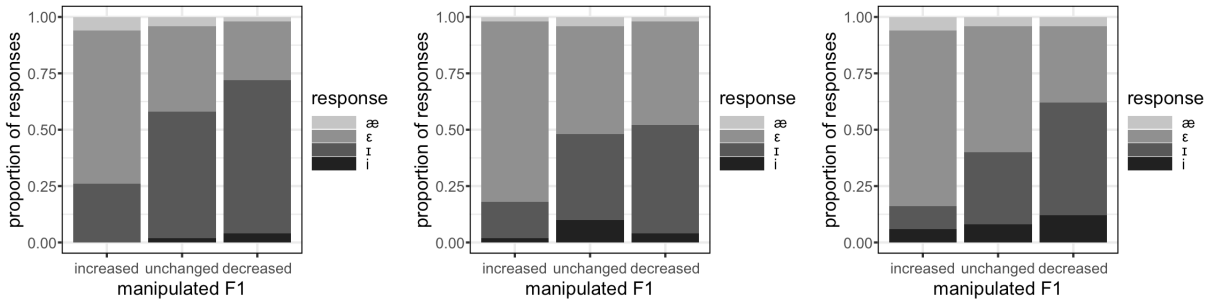
(a) short duration (b) medium duration (c) long duration

Figure 7. Effects of F1 and duration manipulations on identifications of /æ/ stimuli



(a) short duration (b) medium duration (c) long duration

Figure 8. Effects of F1 and duration manipulations on identifications of /ɛ/ stimuli



(a) short duration (b) medium duration (c) long duration

Figure 9. Effects of F1 and duration manipulations on identifications of /ɪ/ stimuli