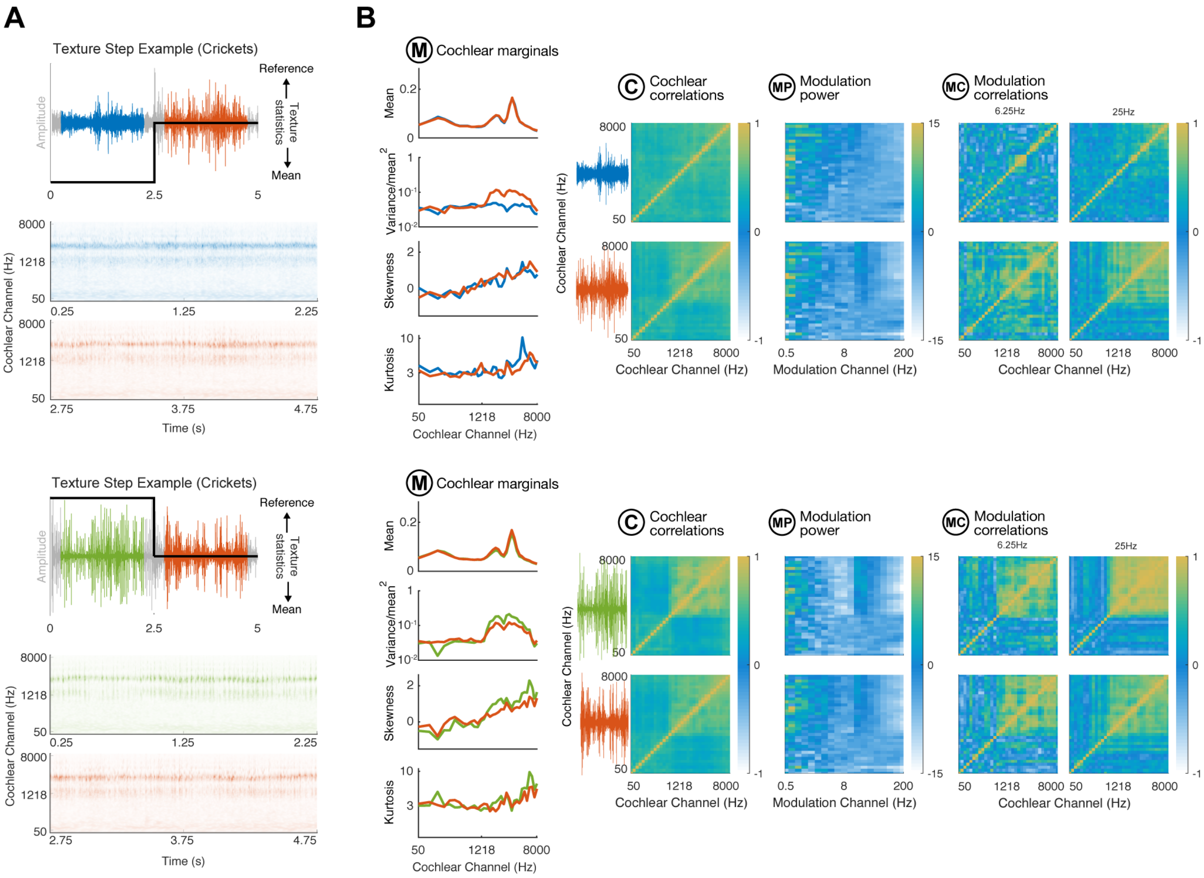

Figure S1 (related to Figure 1)

Statistical variation across synthetic texture steps. **(A)** Waveforms and spectrograms of two example stimuli containing an upward or downward step in texture statistics at 2.5s. **(B)** The statistics of two 2s excerpts from each example stimulus. Here, as in our experimental stimuli, the time-averaged spectrum did not vary across the step (it was always matched to that of the reference texture), hence the similarity across excerpts in the cochlear envelope marginal mean statistics (top panel within the column of marginal statistic plots). All other statistics vary between excerpts, as intended.

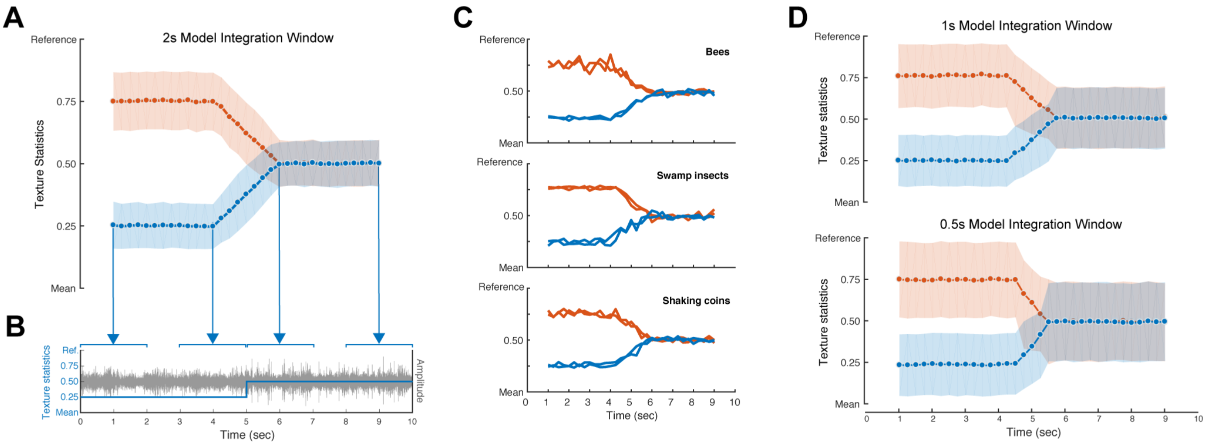

Figure S2 (related to Figure 1)

Texture statistics measured from texture step stimuli. Statistics were measured with a 2s analysis window centered at different positions within the stimulus and then projected on the line between the reference and the mean texture statistics. **(A)** Average measured statistic trajectories for steps moving towards (blue) or away (red) from the reference texture with endpoint statistics midway between the reference and mean. Because the analysis window is long, the measured statistics gradually change over time even though the statistics from which the signal was generated change discretely. However, for measurement windows that do not overlap the step, the statistics are unbiased. Here and in panel D, shaded regions show standard deviation. Stimuli were longer than experimental stimuli in order to more clearly show the transition in statistics across the step. **(B)** Example waveform of one of the stimuli. Blue brackets on top show 2 second analysis window locations corresponding to specific data points on the measured statistic trajectory plot. **(C)** Example statistic trajectories measured from individual stimuli for three reference textures. **(D)** Same as (a), but with 1s (top) and 0.5s (bottom) analysis windows.

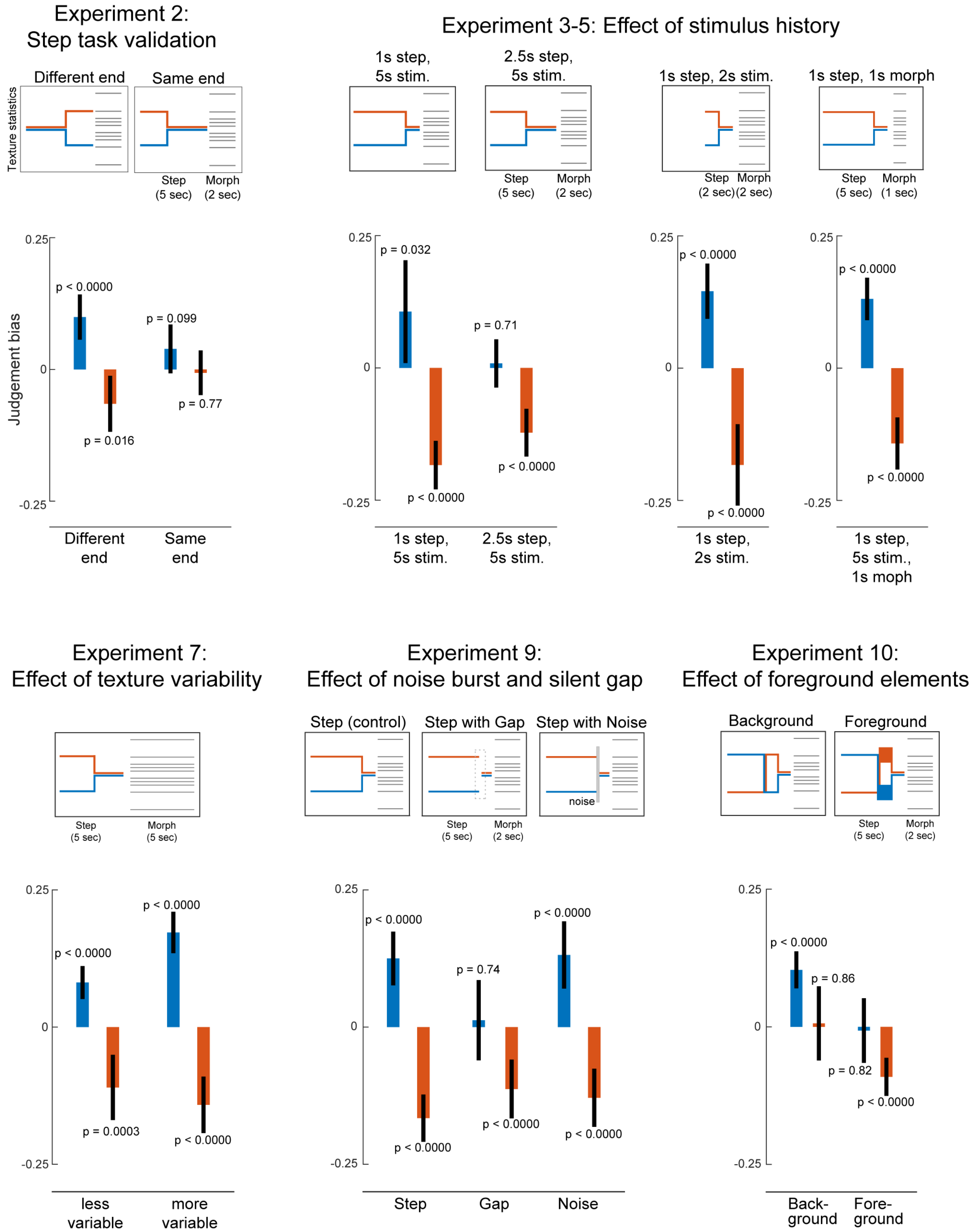

Figure S3 (related to Figures 2, 3, 5, 6, 7)
Discrimination bias for individual step directions for step discrimination experiments (Experiments 2, 3, 4, 5, 7, 9, 10). Bars plot difference between point of subjective equality and 0.5, for the step towards the reference (blue) and away from the reference (red). Error bars show 95% confidence intervals obtained via bootstrap.

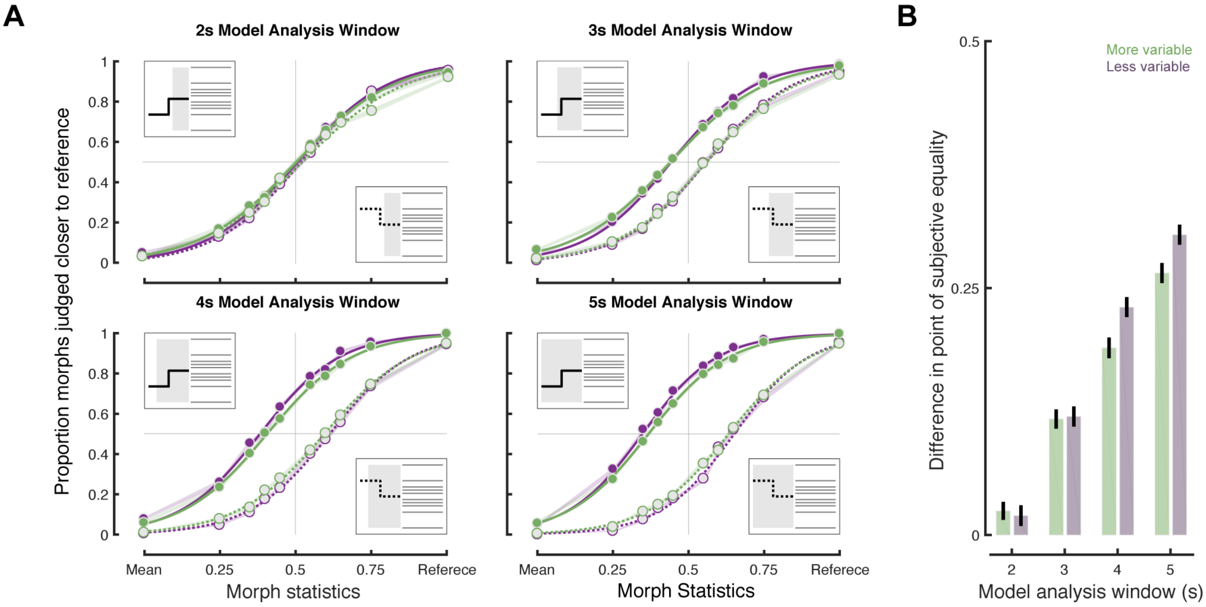

Figure S4 (related to Figure 5)

Results from observer model run on the stimuli from Experiment 7, using four different averaging windows, plotted separately for more and less variable textures. **(A)** Performance of the observer model on a texture step experiment using four different model window sizes (2s, 3s, 4s, and 5s). Shaded region show SEM obtained via bootstrap. Insets show the step direction and morph positions, with the grey region indicating the analysis window extent. **(B)** Bias between paired step conditions for each model analysis window. The bias was similar for the two sets of textures. For the longer analysis windows, the model's bias was in fact slightly greater for the less variable textures (the opposite of what was observed in human listeners). Error bars show 95% confidence intervals on difference (via bootstrap).

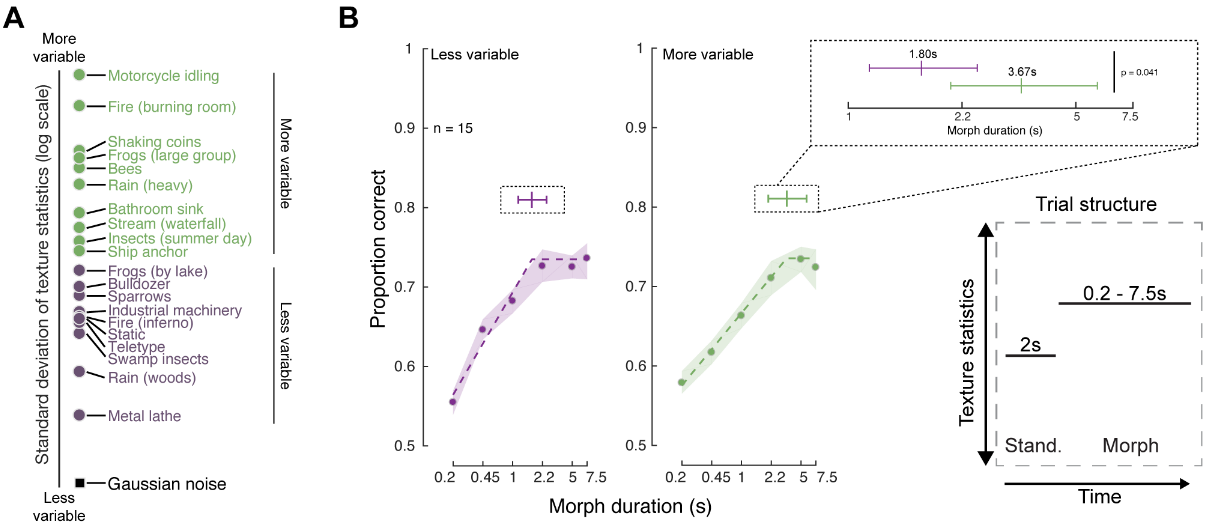

Figure S5 (related to Figure 5)

Influence of texture variability on effect of duration (Re-analysis of Experiment 6). **(A)** Variability in texture statistics measured across 1-s windows for the 20 reference textures used in Experiment 6. The textures in this experiment were not selected to be well differentiated in terms of variability, and so are less separated on this measure than those in Experiment 8; we nonetheless analyzed them in this way as a replication of Experiment 8. **(B)** Discrimination performance vs. duration for less (left) and more (right) variable textures. Bottom right inlay shows schematic of trial structure. Shaded region indicates SEM, obtained via bootstrap. Dashed line shows piecewise linear "elbow" function fit. Top right inset shows expanded view of inflection points of piecewise linear functions fit to data for the more and less homogeneous texture groups (with 95% confidence intervals obtained by bootstrap). The effect of texture variability (longer apparent integration for more variable textures) is qualitatively replicated here. We note that this result additionally replicated in a pilot version of Experiment 6 that used a single morph duration (of 2s) – splitting the textures by variability again produced a later asymptote for more variable textures.

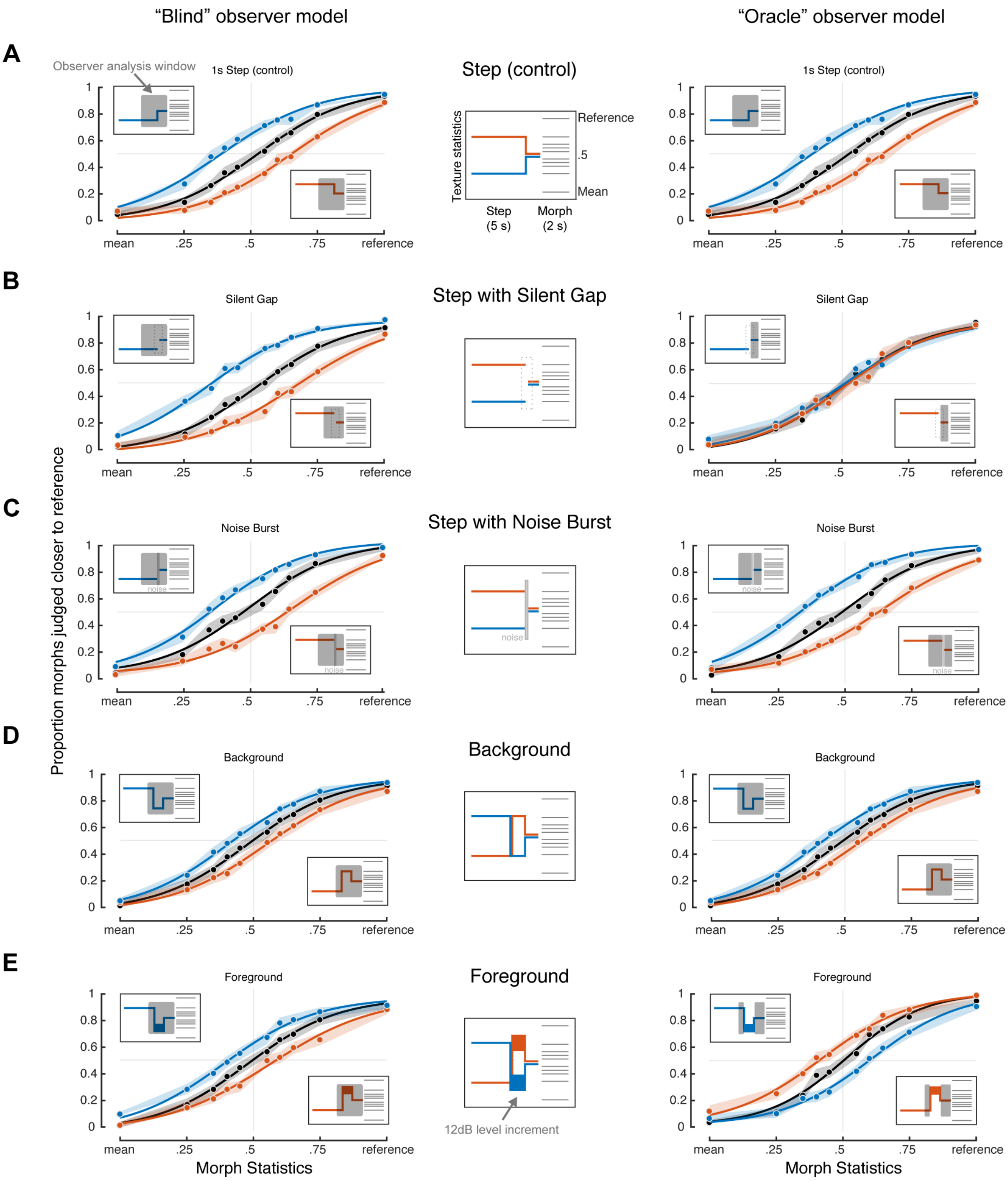

Figure S6 (related to Figures 6 and 7)

Observer model predictions for Experiments 9 and 10. The left column shows results from a "blind" observer model, which measured the statistics of the last 2.5 seconds of the step and the morph and compared the results to that of the reference. The model's analysis window is denoted by the rectangular gray window superimposed on the schematic of the step interval in the insets. The middle column shows the schematic stimuli for each of the 5 conditions. The right column of the figure shows an "oracle" model, which selectively averaged the signal portions hypothesized to underlie human performance, again denoted by rectangular gray shading in the insets. **(A)** The step (control) condition (Experiment 9) included a step in statistics located 1 second from the endpoint. Both models showed the expected bias in the estimated statistics, with the blue curve shifted towards the mean and the red curve shifted towards the reference. Here and elsewhere, shaded region shows SEM obtained via bootstrap and solid line shows logistic function fit. **(B)** The step condition with silent gap (Experiment 9) included a silent gap immediately following the step. The blind observer model was not affected much by the gap, because it averages the signal portions preceding the gap, but the "oracle" model showed no bias because statistics were computed only from the last 800ms of the step interval. **(C)** The step with noise burst condition (Experiment 9) included a 200ms spectrally matched noise burst immediately following the step, replacing the texture during that period. The bias persisted for both models, because the noise burst does not greatly alter the measured statistics irrespective of whether it is included in the average. **(D)** The background condition (Experiment 10) included three step segments. Both observer models yielded a bias towards the position of the second step segment, as intended. **(E)** The foreground condition (Experiment 10) again included three step segments, but the second segment was incremented in level by 12dB. The blind observer model exhibited biases similar to those from the background condition. The oracle model, however, exhibited opposite biases because the 2^{nd} segment was not included in the measurement.

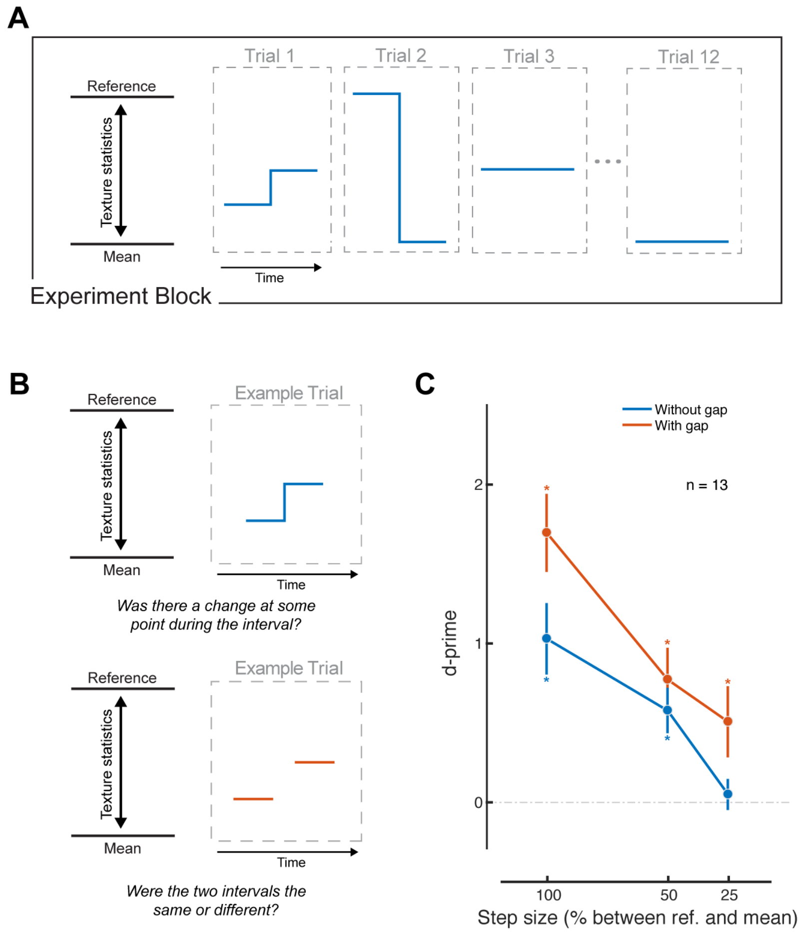

Figure S7 (related to Figures 2 and 6)
Step detection performance with and without a gap (Experiment 11). **(A)** Schematic of block and trial structure used in Experiment 11. Listeners judged whether there was a change in statistics between the first and second halves of the stimulus. A change (step) occurred on half of the trials and varied in magnitude (100%, 50%, and 25% of the distance between the reference and anchor). Obligatory integration should impair change detection relative to a situation where averaging does not occur, and this experiment was intended to test that possibility. **(B)** Schematic of experimental conditions. On the conditions with a gap, the two 1s texture waveforms on either side of the step were separated by 400 ms. **(C)** Step detection results with and without a gap. Asterisks show significant difference from chance (two-tailed t-test, $p < 0.05$). Error bars show SEM obtained via bootstrap. Listeners were at chance when asked to detect the 25% steps used in most of our experiments. Detection improved when the steps were larger, plausibly because such steps begin to introduce audible discontinuities in the stimulus which might either alter integration (as suggested by Experiment 9) or introduce other cues that listeners could use in addition to the output of a statistic estimator. Consistent with this idea, when brief gaps were introduced between the two sides of the step, performance on the otherwise identical task improved. The results are at least consistent with the idea that there is obligatory averaging for continuous textures that is reset by sufficiently salient discontinuities.

Applause	Water running into sink
Motorcycle idling	Bath being drawn
IBM electric typewriter	Stream near small waterfall
Bees	Bulldozer
Frying bacon	Fire – forest inferno
Castanets – rhythmic	Newspaper printing press
Summer day in the south	City room teletype
Fast running river	Static
Wind whistling	Metal lathe
Enthusiastic applause	Frogs – large group
Bathroom sink	Rain in the woods
Shaking coins	Frogs by lake
Industrial machinery	Rain in the woods
Insects in a swamp	Kitchen sink disposal
Bee swarm	Radio static
Ship anchor being raised	Construction site ambience
Sparrows - large excited group	Enthusiastic applause
Horse trotting on cobblestones	Jungle rain
River running over shallows	Rain
Heavy rain falling and dripping	Heavy rain on hard surface
Linotypes	Applause – large crowd
Fire – burning room	Waterfall
Pneumatic drills at road works	Blender
Teletype	Applause – large auditorium
Electric adding machine	Air conditioner

Table S1 (related to Figures 1-7).
50 real-world texture recordings used to create the mean texture.

Motorcycle idling	Bees
Summer day in the south	Bathroom sink
Shaking coins	Industrial machinery
Insects in a swamp	Ship anchor being raised
Sparrows – large excited group	Heavy rain falling and dripping
Fire – burning room	Teletype
Stream near small waterfall	Bulldozer
Fire – forest inferno	Static
Metal lathe	Frogs – large group
Rain in the woods	Frogs by lake

Table S2 (related to Figure 4).
20 real-world texture recordings used in Experiment 6.

Water lapping	Crunching cellophane
Seaside waves	Waterfall
Galloping horses	Mechanical press
Shaking paper	Drumroll
Crunching glass	Small river
Brushing teeth	Frogs
Shaking wood	Swamp noises
Ocean waves	Applause
Raining woods	Pouring coffee beans
Chewing carrot	Heavy rain

Table S3 (related to Figure 5).
10 more (left column) and 10 less (right column) variable real-world texture recordings used in Experiment 8.