Supplementary material for “Pupillometry shows the effort of auditory attention switching”

Daniel R. McCloy, Bonnie K. Lau, Eric Larson, Katherine A. I. Pratt, and Adrian K. C. Lee[a]

Institute for Learning and Brain Sciences, University of Washington, 1715 NE Columbia Rd., Box 357988, Seattle, WA, 98195-7988

March 6, 2017

[a]Author to whom correspondence should be addressed. Electronic mail: akclee@uw.edu

I. REPOSITORY

Code and data for these experiments are available publicly at https://github.com/LABSN-pubs/2017-JASA-pupil-attn-switch

II. STATISTICAL MODEL SPECIFICATION

In both experiments, separate models were constructed for listener sensitivity and for reaction time to target items. Listener sensitivity was modeled with generalized linear mixed-effects regression with a probit link function, using packages `afex`[1] (version 0.16.1) and `lme4`[2] (version 1.1.12) in the `R` statistical computing environment.[3] Reaction time was modeled with linear mixed-effects regression using the same software.

III. EXPERIMENT 1

A. Listener sensitivity

The model for listener sensitivity was constructed to predict probability of button press at each timing slot. Binary fixed-effect predictors specified the trial parameters (maintain/switch, anechoic/reverberant, and talker gender match/mismatch), and also included a categorical indicator variable encoding whether a target, foil, or neither was present in the timing slot. A random intercept was also estimated for each listener. Mathematically, this model is represented as in Equation 1:

$$\Phi^{-1}(y_{ij}) = \beta_0+\beta_1 T_i+\beta_2 F_i+\beta_3 R_i+\beta_4 G_i+\beta_5 A_i+\beta_6 R_i G_i+\beta_7 R_i A_i+\beta_8 G_i A_i+\beta_9 R_i G_i A_i+\beta_{10} T_i R_i+\beta_{11} T_i G_i+\beta_{12} T_i A_i+\beta_{13} T_i R_i G_i+\beta_{14} T_i R_i A_i+\beta_{15} T_i G_i A_i+\beta_{16} T_i R_i G_i A_i+\beta_{17} F_i R_i+\beta_{18} F_i G_i + \beta_{19} F_i A_i + \beta_{20} F_i R_i G_i + \beta_{21} F_i R_i A_i + \beta_{22} F_i G_i A_i + \beta_{23} F_i R_i G_i A_i + S_{0j} \quad (1)$$

where β terms are the coefficients to be estimated, T_i is the binary indicator of target presence, F_i is the binary indicator of foil presence, R_i represents the difference between anechoic and reverberant trials, G_i represents the difference between trials with matching and mismatching talker genders, A_i represents the difference between maintain- and switch-attention trials, and S_{0j} is the random intercept for subject j. β_0 is the grand intercept, β terms subscripted 1 through 9 model response bias, and the remaining β terms model the effect of experimental manipulations on response to target and foil items.

Table I: Coding of indicator variables in the statistical model for Experiment 1.

Term	**Variable**	**Coding**	**Value**
T_i, F_i	`truth`	Treatment	$T_i = 1$ if target present, $F_i = 1$ if foil present; 0 otherwise
A_i	`attn`	Deviation	0.5 if maintain-attention trial, −0.5 if switch-attention trial
R_i	`reverb`	Deviation	0.5 if anechoic trial, −0.5 if reverberant trial
G_i	`gender`	Deviation	0.5 if mismatched-gender trial, −0.5 if matched-gender trial

This model is implemented in `R` as `formula(press ~ truth * reverb * gender * attn + (1|subj))`, where `press` is a binary indicator of whether the listener pressed the response button; `truth` is a treatment-coded factor variable indicating whether a target, foil, or neither was present in the timing slot (with "neither" as baseline); `reverb` is a binary indicator of whether the trial was anechoic or reverberant; `gender` is a binary indicator of whether target and masker voices were both male ("MM") or male target and female masker ("MF"); `attn` is a binary indicator of whether listeners were cued to maintain or switch attention between talkers at the mid-trial gap; and `subj` is an indicator variable for the identity of the listener. Deviation coding was used with all three of the experimental manipulations (`attn`, `reverb`, and `gender`).

Table II: Summary of model of listener sensitivity for Experiment 1. Full-width horizontal rules separate terms modeling baseline response levels, effect of manipulations on response bias, and effect of manipulations on sensitivity. Lighter horizontal rules group together model terms that are tested simultaneously in the likelihood ratio tests. Terms given in SMALL CAPS are the variable names excluded in each likelihood ratio test; terms given in *italics* are the specific values of each variable coded as positive values during modeling (and hence, determine the interpretation of the sign of their corresponding estimate). SE = standard error of coefficient estimates; DF = degrees of freedom; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Model summary							**Likelihood ratio tests**					
		Estimate	SE	Wald z	p	signif.		Model DF	χ^2	χ^2 DF	p	signif.
BASELINE RESPONSE LEVELS												
	(Intercept)	−1.84	0.04	−50.67	<0.001	***						
T_i	*target*	2.99	0.03	95.38	<0.001	***	TRUTH	23	15831.01	2	<0.001	***
F_i	*foil*	0.21	0.04	4.79	<0.001	***						
BIAS TERMS												
R_i	*anech*	−0.06	0.05	−1.33	0.184		REVERB	24	1.77	1	0.184	
G_i	*MF*	−0.03	0.05	−0.58	0.564		GENDER	24	0.33	1	0.564	
A_i	*maint*	−0.12	0.05	−2.57	0.010	*	ATTN	24	6.63	1	0.010	*
R_iG_i	*anech:MF*	0.07	0.10	0.74	0.460		REVERB:GENDER	24	0.55	1	0.460	
R_iA_i	*anech:maint*	−0.12	0.10	−1.24	0.215		REVERB:ATTN	24	1.53	1	0.215	
G_iA_i	*MF:maint*	0.10	0.10	0.99	0.320		GENDER:ATTN	24	0.99	1	0.320	
$R_iG_iA_i$	*anech:MF:maint*	0.10	0.19	0.54	0.592		REVERB:GENDER:ATTN	24	0.29	1	0.592	
SENSITIVITY TERMS												
T_iR_i	*target:anech*	0.19	0.06	3.08	0.002	**	TRUTH:REVERB	23	13.54	2	0.001	**
F_iR_i	*foil:anech*	−0.04	0.09	−0.50	0.618							
T_iG_i	*target:MF*	0.15	0.06	2.43	0.015	*	TRUTH:GENDER	23	19.80	2	<0.001	***
F_iG_i	*foil:MF*	−0.20	0.09	−2.31	0.021	*						
T_iA_i	*target:maint*	0.33	0.06	5.23	<0.001	***	TRUTH:ATTN	23	27.96	2	<0.001	***
F_iA_i	*foil:maint*	0.25	0.09	2.82	0.005	**						
$T_iR_iG_i$	*target:anech:MF*	−0.26	0.12	−2.09	0.036	*	TRUTH:REVERB:GENDER	23	11.26	2	0.004	**
$F_iR_iG_i$	*foil:anech:MF*	0.26	0.17	1.47	0.142							
$T_iR_iA_i$	*target:anech:maint*	−0.01	0.12	−0.09	0.932		TRUTH:REVERB:ATTN	23	5.03	2	0.081	
$F_iR_iA_i$	*foil:anech:maint*	0.35	0.17	1.99	0.047	*						
$T_iG_iA_i$	*target:MF:maint*	0.03	0.12	0.24	0.810		TRUTH:GENDER:ATTN	23	0.06	2	0.971	
$F_iG_iA_i$	*foil:MF:maint*	0.01	0.17	0.07	0.947							
$T_iR_iG_iA_i$	*target:anech:MF:maint*	−0.40	0.25	−1.60	0.109		TRUTH:REVERB:GENDER:ATTN	23	2.64	2	0.267	
$F_iR_iG_iA_i$	*foil:anech:MF:maint*	−0.33	0.35	−0.93	0.351							

B. Reaction time

The model for reaction time was constructed to predict latency of button press at each timing slot. Analysis of only "hit" responses (i.e., button presses occurring between 100 and 1000 ms after the onset of a target) is reported here; an additional analysis that included responses to both targets and foils did not differ in terms of which predictors were significant nor in the direction of the effect for significant predictors (though the magnitude of the estimated effect sizes did vary slightly).

As in the model of listener sensitivity, binary fixed-effect predictors specified the trial parameters (maintain/switch, anechoic/reverberant, and talker gender match/mismatch), but because only hits are analyzed, there was no indicator variable encoding whether a target, foil, or neither was present in the timing slot. Again, a random intercept was estimated for each listener; an error term ϵ_{ij} is also estimated. As measured in this experiment, reaction time is a continuous scalar quantity bounded between 0.1 and 1.0 s. However, examination of the distribution of reaction times resembled a χ^2 distribution (as reaction time measurements often do) and so the response was treated as though continuous and unbounded; hence it was not transformed and no link function was used. Mathematically, the model of reaction time is represented as in Equation 2:

$$(2)\ \ y_{ij} = \beta_0 + \beta_1 R_i + \beta_2 G_i + \beta_3 A_i + \beta_4 R_i G_i + \beta_5 R_i A_i + \beta_6 G_i A_i + \beta_7 R_i G_i A_i + S_{0j} + \epsilon_{ij}$$

This model is implemented in `R` as `formula(reax_time ~ reverb * gender * attn + (1|subj))`, where `reax_time` is reaction time in seconds, and `reverb`, `gender`, `attn`, and `subj` are defined as in the sensitivity model described above.

Table III: Summary of model of reaction time for Experiment 1. Terms given in SMALL CAPS are the variable names excluded in each likelihood ratio test; terms given in *italics* are the specific values of each variable coded as positive values during modeling (and hence, determining the interpretation of the sign of their corresponding estimate). SE = standard error of coefficient estimates; nDF = numerator degrees of freedom; dDF = estimated denominator degrees of freedom; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Model summary					**F tests with Kenward-Roger approximate DF**					
		Estimate	SE	t		nDF	dDF	F	p	signif.
	(Intercept)	0.587	0.020	29.50						
R_i	*anech*	−0.013	0.004	−3.06	REVERB	1	5868.1	9.35	0.002	**
G_i	*MF*	−0.025	0.004	−5.98	GENDER	1	5868.2	35.74	<0.001	***
A_i	*maint*	−0.009	0.004	−2.11	ATTN	1	5868.1	4.45	0.035	*
R_iG_i	*anech:MF*	0.004	0.008	0.49	REVERB:GENDER	1	5868.1	0.24	0.621	
R_iA_i	*anech:maint*	−0.012	0.008	−1.45	REVERB:ATTN	1	5868.0	2.11	0.147	
G_iA_i	*MF:maint*	0.010	0.008	1.18	GENDER:ATTN	1	5868.1	1.39	0.238	
$R_iG_iA_i$	*anech:MF:maint*	0.001	0.016	0.04	REVERB:GENDER:ATTN	1	5868.1	0.00	0.972	

Table IV: Summary of statistical results comparing pupillary response in the different experimental conditions for Experiment 1. Where a single number is given for temporal extent, the cluster comprised only a single sample. *** = $p < 0.001$.

condition	**clusters**	**temporal extent**	***p***	**signif.**
reverberant vs. anechoic	4	3.5 - 3.7 s	0.259	
		3.9 s	0.485	
		4.1 s	0.493	
		4.4 - 4.6 s	0.211	
male-male vs. male-female	2	-0.4 - 0.1 s	0.128	
		3.4 - 3.6 s	0.254	
maintain vs. switch attention	1	1.0 - 5.5 s	<0.001	***

C. Analysis of pupil diameter

Differences in the pupillary response between experimental conditions were calculated using a permutation cluster-level 1-sample t-test,[4] as implemented in `mne-python`. The mean timecourse of pupil size across trials was computed within each condition for each subject, and the within-subject difference between conditions was then calculated. The resulting subject $\times$ time difference matrices were submitted to the permutation function, which simulates (by random sign-flips of the difference matrix) the probability that contiguous cluster(s) at least as large as those attested in the actual data could have arisen by chance. For all tests, the t-statistic threshold was approximately 2.13 (based on a two-sided hypothesis test with 2.5% lower-tail probability and 1 degree of freedom). All possible permutations were performed (for 16 subjects, this is 2^{15} or 32768). Statistics for the differences between conditions for the three experimental manipulations are shown in Table IV.

D. Post-hoc analyses

1. *Foil response rate by slot*

To address the concern that listeners might have attempted to monitor both streams, and especially that they might do so differently in maintain- versus switch-attention trials, the rate of listener response to foil items was examined separately for each timing slot. Post-hoc analysis comparing logit-transformed foil response rates in maintain- versus switch-attention trials, performed separately for each timing slot, showed no significant differences (paired t-tests: p=0.97, 0.96, 0.70, 0.49 for slots 1–4 respectively; Bonferroni-corrected significance level 0.0125). The logit transformation was necessary because foil response rates are proportions bounded between 0 and 1, making t-tests on the untransformed rates inappropriate.

2. *Reaction time by slot*

Post-hoc analysis of reaction time for the reverberation contrast showed no significant differences by response slot (Welch's independent t-tests: p=0.41, 0.010, 0.037, 0.043 for slots 1–4 respectively). For the talker gender (mis)match contrast, post-hoc analysis of reaction time showed a significant difference only for slot 3 (Welch's independent t-tests: p=0.046, 0.063, 0.00018, 0.0065 for slots 1–4 respectively). For the maintain- versus switch-attention contrast, post-hoc analysis of reaction time also showed a significant difference only for slot 3 (Welch's independent t-tests: p=0.75, 0.53, 0.001, 0.035 for slots 1–4 respectively). Bonferroni-corrected significance level for combined post-hoc analyses = 0.00417.

3. Gaze direction

Measures of pupil size can be affected by as much as 10% when gaze direction is oblique to the EyeLink camera. To ensure this did not affect our measures of pupil size, we checked the distribution of gaze angles relative to the fixation cross. Plots of the distribution of gaze angles for each subject are shown in Figure 1; overall percentages of fixations relative to various threshold values are shown in Table V.

Table V: Distribution of gaze direction relative to fixation cross.

Threshold	**Fixations below threshold**
2°	85.8%
5°	98.5%
10°	99.9%

IV. EXPERIMENT 2

A. Listener sensitivity

As in Experiment 1, the model for listener sensitivity was constructed to predict probability of button press at each timing slot. Binary fixed-effect predictors specified the trial parameters (maintain/switch, 10/20 channel vocoding, and 200/600 ms mid-trial switch gap duration), and also included a categorical indicator variable encoding whether a target, foil, or neither was present in the timing slot. A random intercept was also estimated for each listener. Mathematically, this model is represented as in Equation 3:

$$(3)\quad \Phi^{-1}(y_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 F_i + \beta_3 V_i + \beta_4 G_i + \beta_5 A_i + \beta_6 V_i G_i + \beta_7 V_i A_i + \beta_8 G_i A_i + \beta_9 V_i G_i A_i + \beta_{10} T_i V_i + \beta_{11} T_i G_i + \beta_{12} T_i A_i + \beta_{13} T_i V_i G_i + \beta_{14} T_i V_i A_i + \beta_{15} T_i G_i A_i + \beta_{16} T_i V_i G_i A_i + \beta_{17} F_i V_i +$$

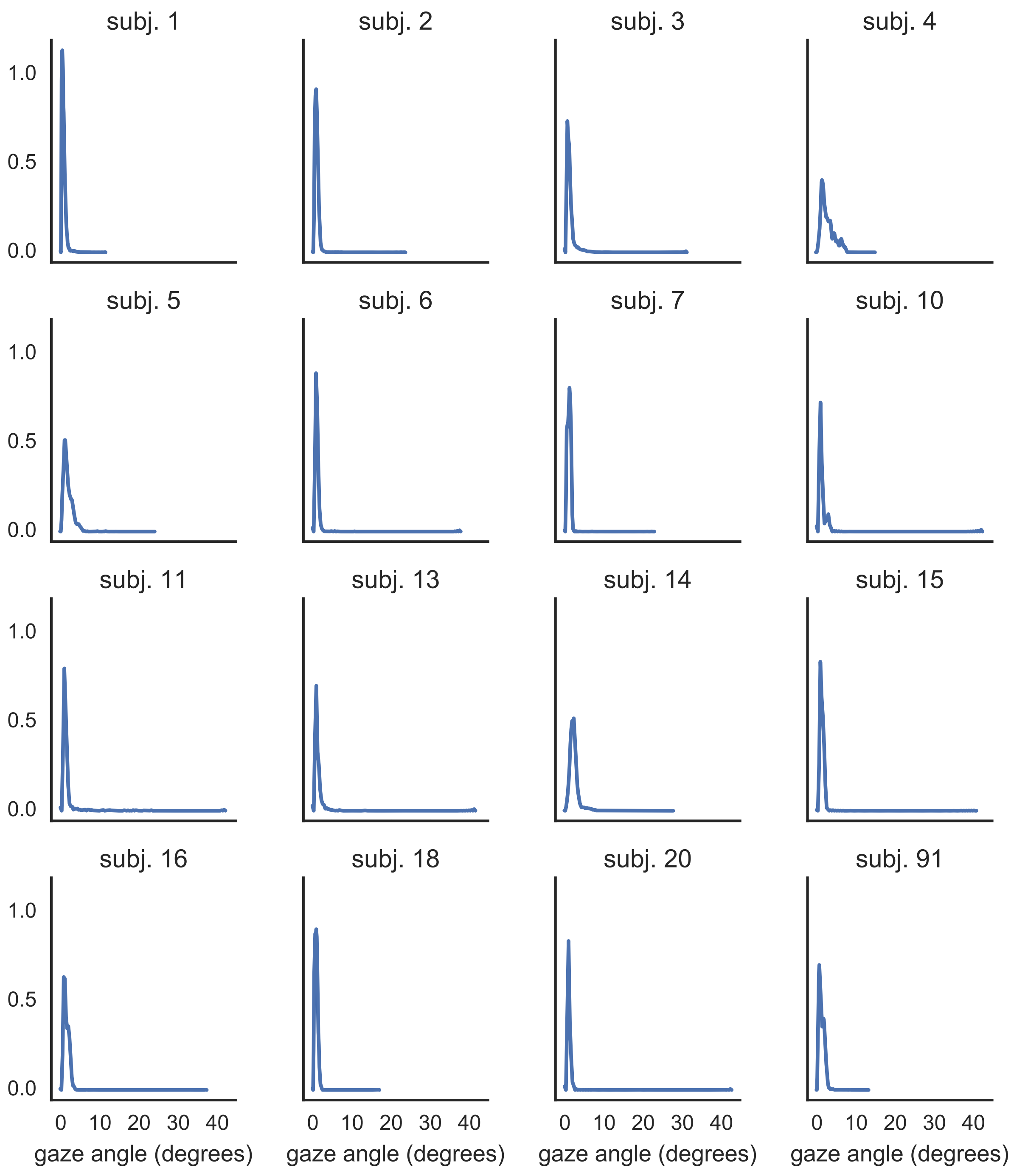

Figure 1: Distribution of gaze angles (relative to fixation cross) for each subject.

$$\beta_{18}F_iG_i + \beta_{19}F_iA_i + \beta_{20}F_iV_iG_i + \beta_{21}F_iV_iA_i + \beta_{22}F_iG_iA_i + \beta_{23}F_iV_iG_iA_i + S_{0j}$$

where β terms are the coefficients to be estimated, T_i is the binary indicator of target presence, F_i is the binary indicator of foil presence, V_i represents the difference between 20- and 10-channel vocoder trials, G_i represents the difference between trials with matching and mismatching talker genders, A_i represents the difference between maintain- and switch-attention trials, and S_{0j} is the random intercept for subject j. β_0 is the grand intercept, β terms subscripted 1 through 9 model response bias, and the remaining β terms model the effect of experimental manipulations on response to target and foil items.

Table VI: Coding of indicator variables in the statistical model for Experiment 2.

Term	**Variable**	**Coding**	**Value**
T_i, F_i	`truth`	Treatment	$T_i = 1$ if target present, $F_i = 1$ if foil present; 0 otherwise
A_i	`attn`	Deviation	0.5 if maintain-attention trial, -0.5 if switch-attention trial
V_i	`voc_chan`	Deviation	0.5 if 20-channel trial, -0.5 if 10-channel trial
G_i	`gap_len`	Deviation	0.5 if long-gap trial, -0.5 if short-gap trial

This model is implemented in R as `formula(press ~ truth * voc_chan * gap_len * attn + (1|subj))`, where `press` is a binary indicator of whether the listener pressed the response button; `truth` is a treatment-coded factor variable indicating whether a target, foil, or neither was present in the timing slot (with “neither” as baseline); `voc_chan` is a binary indicator of whether the trial was processed with 10- or 20-channel noise vocoding; `gap_len` is a binary indicator of mid-trial gap duration (200 or 600 ms); `attn` is a binary indicator of whether listeners were cued to maintain or switch attention between talkers at the mid-trial gap; and `subj` is an indicator variable for the identity of the listener. Deviation coding was used with all three of the experimental manipulations (`attn`, `voc_chan`, and `gap_len`).

Table VII: Summary of model of listener sensitivity for Experiment 2. Full-width horizontal rules separate terms modeling baseline response levels, effect of manipulations on response bias, and effect of manipulations on sensitivity. Lighter horizontal rules group together model terms that are tested simultaneously in the likelihood ratio tests. Terms given in SMALL CAPS are the variable names excluded in each likelihood ratio test; terms given in *italics* are the specific values of each variable coded as positive values during modeling (and hence, determine the interpretation of the sign of their corresponding estimate). SE = standard error of coefficient estimates; DF = degrees of freedom; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Model summary							**Likelihood ratio tests**					
		Estimate	SE	Wald z	p	signif.		Model DF	χ^2	χ^2 DF	p	signif.
BASELINE RESPONSE LEVELS												
	(Intercept)	−1.36	0.03	−43.75	<0.001	***						
T_i	*target*	1.79	0.02	75.78	<0.001	***	TRUTH	23	6671.88	2	<0.001	***
F_i	*foil*	0.70	0.03	23.31	<0.001	***						
BIAS TERMS												
V_i	*20*	0.02	0.04	0.64	0.521		VOC__CHAN	24	0.41	1	0.521	
G_i	*long*	0.05	0.04	1.42	0.154		GAP__LEN	24	2.03	1	0.154	
A_i	*maint*	0.06	0.04	1.59	0.112		ATTN	24	2.53	1	0.112	
V_iG_i	*20:long*	−0.08	0.07	−1.12	0.263		VOC__CHAN:GAP__LEN	24	1.25	1	0.263	
V_iA_i	*20:maint*	−0.00	0.07	−0.01	0.995		VOC__CHAN:ATTN	24	0.00	1	0.994	
G_iA_i	*long:maint*	−0.06	0.07	−0.84	0.400		GAP__LEN:ATTN	24	0.71	1	0.399	
$V_iG_iA_i$	*20:long:maint*	0.14	0.14	1.03	0.304		VOC__CHAN:GAP__LEN:ATTN	24	1.06	1	0.304	
SENSITIVITY TERMS												
T_iV_i	*target:20*	0.19	0.05	4.09	<0.001	***	TRUTH:VOC__CHAN	23	18.20	2	<0.001	***
F_iV_i	*foil:20*	0.04	0.06	0.69	0.489							
T_iG_i	*target:long*	−0.35	0.05	−7.51	<0.001	***	TRUTH:GAP__LEN	23	254.08	2	<0.001	***
F_iG_i	*foil:long*	0.56	0.06	9.24	<0.001	***						
T_iA_i	*target:maint*	0.05	0.05	1.07	0.284		TRUTH:ATTN	23	12.33	2	0.002	**
F_iA_i	*foil:maint*	−0.15	0.06	−2.54	0.011	*						
$T_iV_iG_i$	*target:20:long*	0.16	0.09	1.66	0.097		TRUTH:VOC__CHAN:GAP__LEN	23	11.38	2	0.003	**
$F_iV_iG_i$	*foil:20:long*	−0.23	0.12	−1.92	0.055							
$T_iV_iA_i$	*target:20:maint*	−0.03	0.09	−0.36	0.718		TRUTH:VOC__CHAN:ATTN	23	1.87	2	0.392	
$F_iV_iA_i$	*foil:20:maint*	0.12	0.12	1.03	0.303							
$T_iG_iA_i$	*target:long:maint*	−0.10	0.09	−1.03	0.305		TRUTH:GAP__LEN:ATTN	23	15.55	2	<0.001	***
$F_iG_iA_i$	*foil:long:maint*	0.36	0.12	2.98	0.003	**						
$T_iV_iG_iA_i$	*target:20:long:maint*	−0.05	0.19	−0.29	0.771		TRUTH:VOC__CHAN:GAP__LEN:ATTN	23	4.07	2	0.131	
$F_iV_iG_iA_i$	*foil:20:long:maint*	−0.46	0.24	−1.92	0.055							

B. Reaction time

As in Experiment 1, the model for reaction time was constructed to predict latency of button press at each timing slot. Again, only "hit" responses were analyzed; an additional analysis that included responses to both targets and foils did not differ in terms of which predictors were significant nor in the direction of the effect for significant predictors (though the magnitude of the estimated effect sizes did vary slightly).

As in the model of listener sensitivity for Experiment 2, binary fixed-effect predictors specified the trial parameters (maintain/switch, 10/20 channel vocoding, and 200/600 ms mid-trial switch gap duration), but because only hits are analyzed, there was no indicator variable encoding whether a target, foil, or neither was present in the timing slot. Again, a random intercept for each listener and an error term were estimated, and the response was treated as continuous and unbounded. Mathematically, the model of reaction time is represented as in Equation 4:

$$(4)\quad y_{ij} = \beta_0 + \beta_1 V_i + \beta_2 G_i + \beta_3 A_i + \beta_4 V_i G_i + \beta_5 V_i A_i + \beta_6 G_i A_i + \beta_7 V_i G_i A_i + S_{0j} + \epsilon_{ij}$$

This model is implemented in `R` as `formula(reax_time ~ voc_chan * gap_len * attn + (1|subj))`, where `reax_time` is reaction time in seconds, and `voc_chan`, `gap_len`, `attn`, and `subj` are defined as in the sensitivity model described above.

Table VIII: Summary of model of reaction time for Experiment 2. Terms given in SMALL CAPS are the variable names excluded in each likelihood ratio test; terms given in *italics* are the specific values of each variable coded as positive values during modeling (and hence, determining the interpretation of the sign of their corresponding estimate). SE = standard error of coefficient estimates; nDF = numerator degrees of freedom; dDF = estimated denominator degrees of freedom; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Model summary					**F tests with Kenward-Roger approximate DF**					
		Estimate	SE	t		nDF	dDF	F	p	signif.
	(Intercept)	0.653	0.019	34.45						
V_i	*20*	−0.035	0.007	−4.67	VOC__CHAN	1	4605.0	21.79	<0.001	***
G_i	*long*	−0.066	0.007	−8.80	GAP__LEN	1	4606.9	77.52	<0.001	***
A_i	*maint*	−0.014	0.007	−1.93	ATTN	1	4605.1	3.73	0.054	
V_iG_i	*20:long*	−0.044	0.015	−2.93	VOC__CHAN:GAP__LEN	1	4604.4	8.57	0.003	**
V_iA_i	*20:maint*	0.008	0.015	0.52	VOC__CHAN:ATTN	1	4604.3	0.27	0.602	
G_iA_i	*long:maint*	0.007	0.015	0.44	GAP__LEN:ATTN	1	4604.8	0.19	0.661	
$V_iG_iA_i$	*20:long:maint*	−0.010	0.030	−0.33	VOC__CHAN:GAP__LEN:ATTN	1	4604.2	0.11	0.744	

Table IX: Summary of statistical results comparing pupillary response in the different experimental conditions for Experiment 2. ** = $p < 0.01$, *** = $p < 0.001$.

condition	clusters	temporal extent	*p*	signif.
reverberant vs. anechoic	2	2.6 - 3.2 s	0.074	
		3.9 - 5.0 s	0.004	**
10- vs. 20-channel vocoding	1	0.7 - 0.9 s	0.161	
maintain vs. switch attention	1	0.9 - 5.6 s	<0.001	***

C. Analysis of pupil diameter

As in Experiment 1, differences in the pupillary response between experimental conditions were calculated using a permutation cluster-level 1-sample t-test. The t-statistic threshold was again approximately 2.13, and all possible permutations were performed. Statistics for the differences between conditions for the three experimental manipulations are shown in Table IX.

D. Post-hoc analyses

1. Reaction time by slot

Post-hoc tests of reaction time difference within each timing slot between maintain- and switch-attention trials showed a significant difference in reaction time localized to slot 3 (the immediately post-gap slot), with longer reaction times in switch-attention trials (Welch's independent t-tests: p=0.585, 0.378, 0.001, 0.916 for slots 1–4 respectively). For the spectral degradation contrast, there was a significant difference localized to slot 1 (Welch's independent t-tests: p=0.002, 0.012, 0.230, 0.891 for slots 1–4 respectively). The gap length manipulation showed significant differences between conditions by slot (Welch's independent t-tests: p<0.001, p=0.060, p<0.001, and p<0.001 for slots 1–4 respectively). Significantly faster reaction times were seen in the long-gap trials for slot 3 (155 ms faster on average) and slot

4 (135 ms faster on average), and significantly *slower* reaction times in the long-gap trials for slot 1 (261 ms slower on average). Bonferroni-corrected significance level for combined post-hoc analyses = 0.00417.

2. Gaze direction

As in Experiment 1, we checked the distribution of gaze angles relative to the fixation cross. Plots of the distribution of gaze angles for each subject are shown in Figure 2; overall percentages of fixations relative to various threshold values are shown in Table X.

Table X: Distribution of gaze direction relative to fixation cross.

Threshold	**Fixations below threshold**
2°	74.9%
5°	99.0%
10°	99.9%

V. REFERENCES

[1] Singmann, H., Bolker, B., Westfall, J., and Aust, F. (**2016**). “Afex: Analysis of factorial experiments,” R package version 0.16-1. Retrieved from https://CRAN.R-project.org/

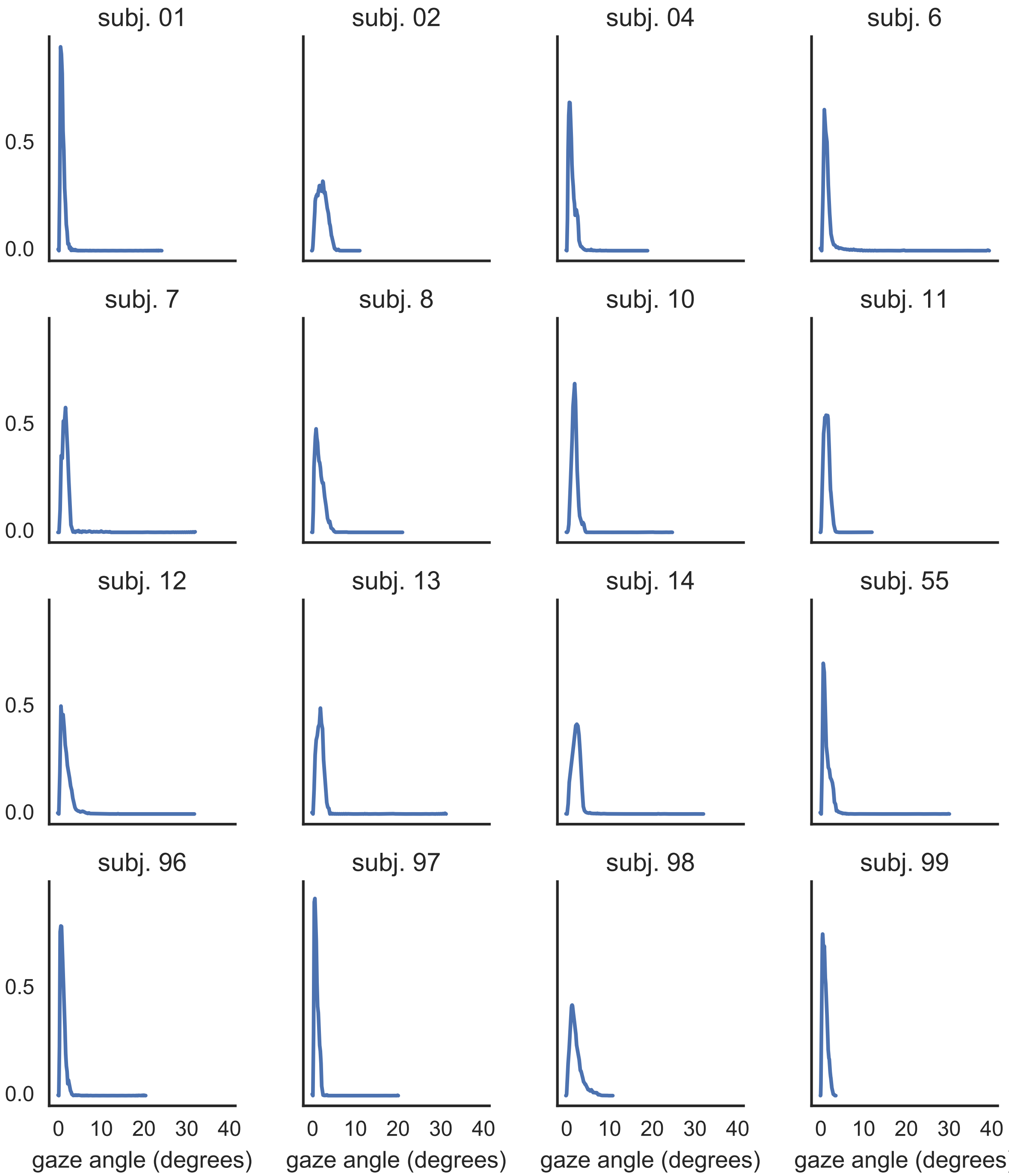

Figure 2: Distribution of gaze angles (relative to fixation cross) for each subject.

package=afex

[2] Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," Journal of Statistical Software, **67**, 1–48. doi:10.18637/jss.v067.i01

[3] R Development Core Team (**2016**). "R: A language and environment for statistical computing," version 3.3.1. Retrieved from http://www.R-project.org/

[4] Maris, E., and Oostenveld, R. (**2007**). "Nonparametric statistical testing of EEG- and MEG-data," J. Neurosci. Meth., **164**, 177–190. doi:10.1016/j.jneumeth.2007.03.024