Supplemental Material S2. LMER summary table, reporting the reference level of the factor block, the model estimate, standard error of the estimate, degrees of freedom for the fixed factor, and the *t*- and *p*-values.

Fixed Effect	Block Ref. Level	Estimate	Std. Error	*df*	*t*	Pr(>|*t*|)
Intercept	Block 1	32.998	3.941	72.572	8.373	< .001
Block 2	Block 1	9.897	1.787	1715.129	5.538	< .001***
Block 3	Block 2	2.722	1.784	1714.183	1.526	.127
Block 4	Block 3	5.151	1.785	1714.643	2.885	.004**

Model Syntax: lmer(accuracy ~ (1|participant2) + (1|sentence) + block)

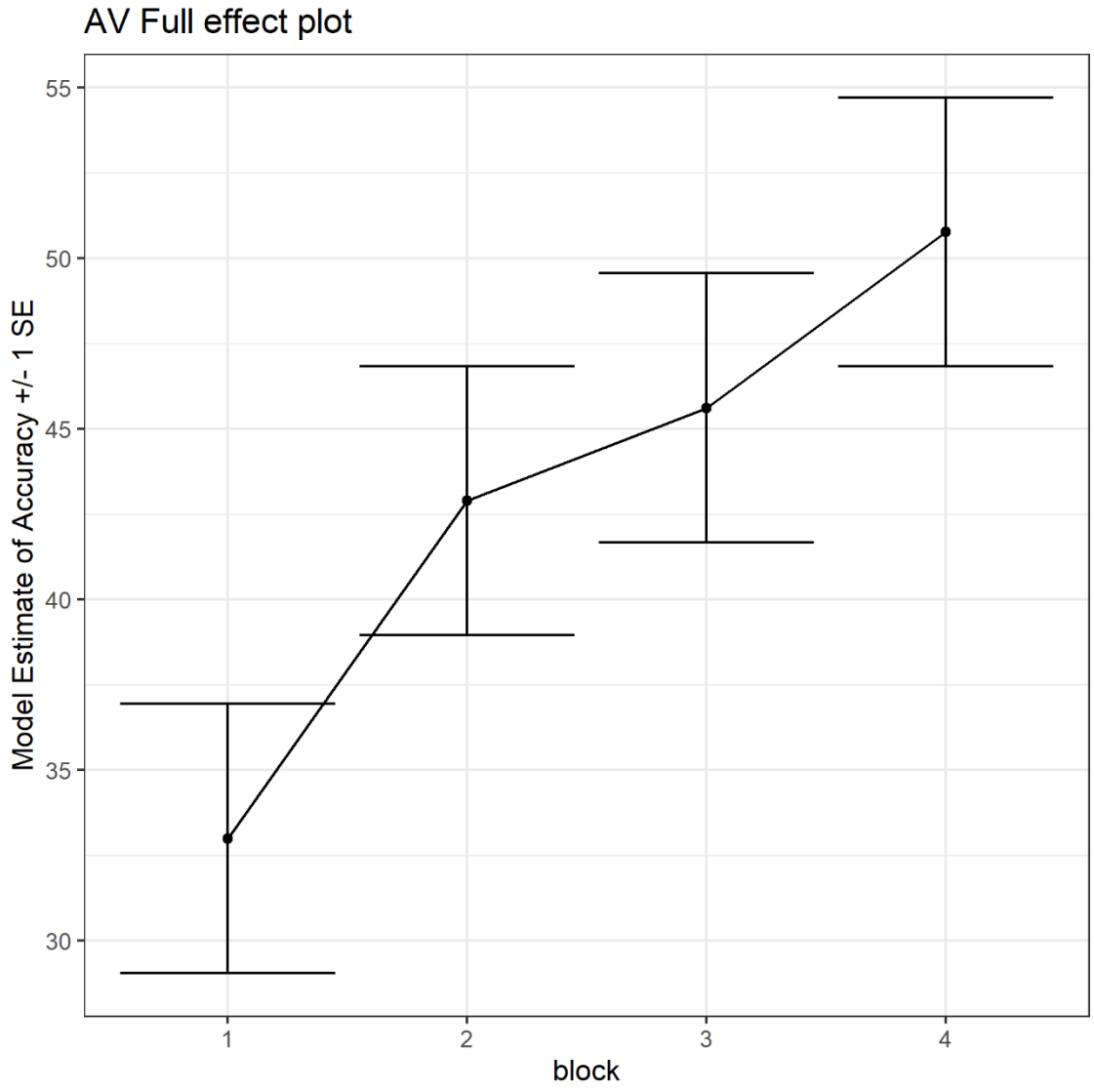

Figure S1. Model estimates of percentage correctly reported key words across the four blocks in condition AV Full, error bars represent one standard error.

Supplementary Analysis 2 - All conditions by block

In the main paper, stepwise model building indicated that the main effect of trial better captured variance across the five conditions. However, as outlined in pre-registration number ##41527 "Transcribing distorted audiovisual speech," we proposed a by-blocks analysis of adaptation. This analysis is therefore detailed here.

To analyse all five conditions, we followed the same procedure as the analysis for the AV Full condition, while including testing condition (factor-coded) as an additional main effect, and the interaction between condition and block. The maximal model upon which we conducted the backward step-wise model comparison therefore included by-items, by-participants random intercepts, a random intercept for participant nested within condition, the main effects of block and condition, and the two-way interaction between block and condition. Random slopes were excluded from the analysis, as their inclusion resulted in issues of singular model fit. The backwards stepwise model selection indicated that the inclusion of the simple by-participants random effect did not improve model fit ($p = 1$). As a result, the final model included a random by-items random intercept, a random-intercept for participant nested within condition, the main effects of block and condition, and the two-way interaction between condition and block (see Supplemental Material S3 for full model syntax and model summary).

To assess the likelihood of the alternative hypothesis – that different conditions would elicit different levels of adaptation – we performed model comparison between a model including the interaction (H_1, BIC = 81002) against a null model only including the main effects (H_0, BIC = 80952), and generated the inverse of BF_{01} ($1.388e^{-11}$). The inverse of BF_{01} was therefore 72004899337, vastly exceeding Rafferty's (1995) threshold for strong evidence (> 150) in favour of the alternative hypothesis. This reflects the floor performance seen in the AV Blocked, AV Eyes, and AV Still conditions relative to the AV Full, and AV Mouth conditions.

The outcomes of the linear model included a main effect of condition (denoted by the intercept when each condition is taken as a reference level, in Supplemental Material S3), wherein participants performance in the AV Full ($t = 14.302$, $p < .0001$), AV Mouth ($t = 14.234$, $p < .0001$) differed significantly from 0, whilst participants in the AV Block ($t = 1.767$, $p = .079$), AV Eyes ($t = 1.711$, $p = .088$), and AV Still ($t = 1.431$, $p = .154$) condition did not, suggesting low initial accuracy in these conditions. Accuracy in the AV Full and AV Full condition did not differ ($t = -0.057$, $p = .954$), reflecting similar initial performance in these conditions (see Figure S2 for greater detail). The significant interaction with block showed that participants in the AV Full condition significantly increased by between blocks 1 and 2 ($t = 6.783$, $p < .0001$), and blocks 3 and 4 ($t = 3.764$, $p = .0002$), however performance did not differ between blocks 2 and 3 ($t = 1.686$, $p = .092$) (see Figure S2 for greater detail). The AV Mouth condition had similar performance, with significant increases in performance between blocks 1 and 2 ($t = 7.920$, $p < .0001$), 2 and 3 ($t = 2.077$, $p = .034$), and 3 and 4 ($t = 2.157$, $p = .031$). The improvements in performance between blocks, however, did not significantly differ between the AV Full and AV Mouth conditions between blocks 1 and 2 ($t = -1.079$, $p = .281$), 2 and 3 ($t = -0.276$, $p = .782$), and 3 and 4 ($t = 1.136$, $p = .256$). The AV Block, AV still, and AV Eyes conditions did not show improvement across blocks, nor did they significantly differ from one another ($ts < 2$, $p > .05$). The AV Mouth condition showed significantly greater increases in performance over the AV Block, AV Eyes and AV Still conditions between each block, whilst the AV Full condition did not show significantly greater increases in performance between blocks 2 and 3 over these conditions, reflecting the flat performance between these blocks.

To summarise, only the AV Mouth and AV Full conditions showed adaptation across blocks, though they did not statistically differ from one another. The AV Eyes, AV Full, and AV Still condition did not show adaptation, and all performed similarly. Taken together, this suggests that participants require being able to see speech motor activity to be able to adapt to distorted audiovisual speech in an online testing environment.

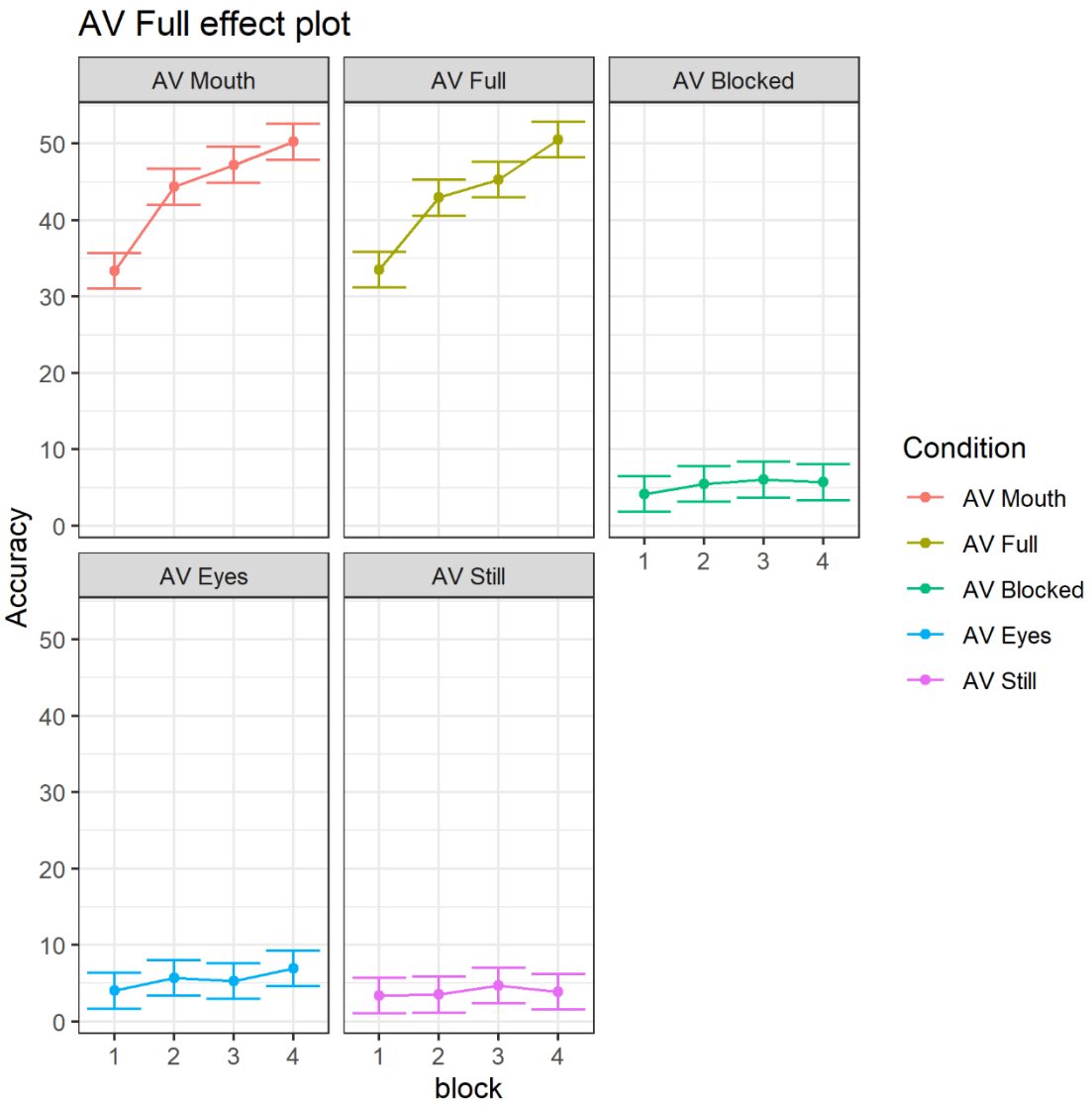

Figure S2. Model estimates of percentage correctly reported key words across the four blocks in all five conditions, error bars represent one standard error.