¹ Listeners adjust their prior expectations as they adapt to speech of an unfamiliar talker

² Maryann Tan[1,2], T Florian Jaeger[2,3], & YOUR OTHER CO-AUTHOR[2]

³ [1] Centre for Research on Bilingualism, University of Stockholm

⁴ [2] Brain and Cognitive Sciences, University of Rochester

⁵ [3] Computer Science, University of Rochester

⁶ Author Note

⁸ Correspondence concerning this article should be addressed to Maryann Tan, YOUR

⁹ ADDRESS. E-mail: maryann.tan@biling.su.se

1   Abstract

YOUR ABSTRACT GOES HERE. All data and code for this study are shared via OSF,

including the R markdown document that this article is generated from, and an R library that

implements the models we present.

Keywords: speech perception; perceptual adaptation; distributional learning; ...

Word count: X

2 Listeners adjust their prior expectations as they adapt to speech of an unfamiliar talker

TO-DO

2.1 Highest priority

- MARYANN
  - Continue describing Experiment 2

2.1.1 Priority

- MARYANN
  - Fix spread_draws bug

2.2 To do later

- Everyone: Eat ice-cream and perhaps have a beer.

1   Introduction

Talkers who share a common language vary in the way they pronounce its linguistic categories. Yet, listeners of the same language background typically cope with such variation without much trouble. In scenarios where a talker produces those categories in an unexpected and unfamiliar way, comprehending their speech may pose a real challenge. However, brief exposure to the talker's accent (sometimes just minutes) can be sufficient for the listener to overcome any initial comprehension difficulty (e.g. Bradlow & Bent, 2008; Clarke & Garrett, 2004; X. Xie, Liu, & Jaeger, 2021; X. Xie et al., 2018). This adaptive skill is in a sense, trivial for any expert language user but becomes complex when considered from the angle of acoustic-cue-to-linguistic-category mappings. Since talkers differ in countless ways and each listening occasion is different in circumstance, there is not a single set of cues that can be definitively mapped to each linguistic category. Listeners instead have to contend with many possible cue-to-category mappings and infer the intended category of the talker. How listeners achieve prompt and accurate comprehension of speech in spite of this variability remains the overarching aim of speech perception research.

Researchers have been exploring the hypothesis that listeners solve this perceptual problem by exploiting their knowledge gained from experience with different talkers. This knowledge is often implicit and context contingent since listeners are sensitive to both social and environmental cues (e.g. age, sex, group identity, native language etc.) that are relevant for optimal speech perception. Impressively, shifts in perception can be induced implicitly through subtle cues such as the presence of cultural artefacts that hint at talker provenance, (Hay & Drager, 2010) and explicitly such as when the listener is instructed to imagine a talker as a man or a woman (Johnson, Strand, & D'Imperio, 1999). While these and other related effects of exposure-induced changes speak to the malleability of human perception, it remains unclear how human perceptual systems strike the balance between stability and flexibility.

One possibility is that listeners continuously update their implicit knowledge with each talker encounter by integrating prior knowledge of cue-to-category distributions with the statistics of the current talker's productions, leading to changes in representations which affect listener categorisation behaviour. Broadly speaking, many theoretical accounts would agree with this

assertion. Connectionist (McClelland & Elman 1986; Luce & Pisoni, 1998), and Bayesian models of spoken word recognition (Norris & McQueen, 2008) and adaptation (Kleinschmidt & Jaeger, 2015) are generative systems that abstract the frequency of input. Even exemplar models of speech perception (Goldinger 1996, 1998; Johnson, 1997; Pierrehumbert 2001) which encode high fidelity memories of speaker-specific phonetic detail converge to a level of generalisation due to effects of token frequency (**Pierrehumbert2003?**; **DragerKirtley2016?**).

At the level of acoustic-phonetic input, listeners' implicit knowledge refer to the way relevant acoustic cues that distinguish phonological categories are distributed across talkers within a linguistic system. Talkers of US-English, for instance, distinguish the /d/-/t/ contrasts primarily through the voice-onset-time (VOT) acoustic cue. Given its relevance for telling word pairs such as "din" and "tin" apart, a distributional learning hypothesis would posit that listeners learn the distribution of VOT cues when talkers produce those stop consonant contrasts in word contexts. Earliest evidence for listener sensitivity to individual talker statistics in the domain of stop consonants come from studies such as Allen & Miller (2004, also Theodore & Miller, 2010) but more recent studies that formalise the problem of speech perception as rational inference have shown that listeners' behavioural responses are probabilistic function of the exposure talker's statistics (Clayards, Tanenhaus, Aslin, & Jacobs, 2008a; Kleinschmidt & Jaeger, 2016; and Theodore & Monto, 2019).

Clayards et al. (2008a) for instance found that listeners responded with greater uncertainty after they were exposed to VOT distributions for a "beach-peach" contrast that had wider variances as compared to another group who had heard the same contrasts with narrower variances. Across both wide and narrow conditions, the mean values of the voiced and voiceless categories were kept constant and set at values that were close to the expected means for /b/ and /p/ in US English. The study was one of the first to demonstrate that at least in the context of an experiment, listeners categorisation behaviour was a function of the variance of the exposure talker's cue distributions – listeners who were exposed to a wide distribution of VOTs showed greater uncertainty in their perception of the stimuli, exhibiting a flatter categorisation function on average, compared to listeners who were exposed to a narrow distribution.

In a later study Kleinschmidt and Jaeger (2016) tested listener response to talker statistics

by shifting the means of the voiced and voiceless categories between conditions. Specifically, the mean values for /b/ and /p/ were shifted rightwards by several magnitudes, as well as leftwards, from the expected mean values of a typical American English talker while the category variances remained identical and the distance between the category means were kept constant. With this manipulation of means they were able to investigate how inclined listeners are to adapt their categorisation behaviors when the statistics of the exposure talker were shifted beyond the bounds of a typical talker.

In all exposure conditions, listeners on average adapted to the exposure talker by shifting their categorization towards the boundary implied by the exposure distribution. However, in all conditions, listener categorization fell short of the predicted ideal categorization boundary. This difference between the observed and predicted categorization functions was larger, the greater the magnitude of the shift from the typical talker's distribution, suggesting adaptation was constrained by listeners' prior experience.

The study we report here builds on the pioneering work of Clayards et al. (2008a) and Kleinschmidt and Jaeger (2016) with the aim to shed more light on the role of prior implicit knowledge on adaptation to an unfamiliar talker.

Specifically, while K&J16 demonstrated how prior beliefs of listeners can be inferred computationally from post-exposure categorisation, their experiment was not designed to capture listener categorisation data before exposure to a novel talker. Nor did they run intermittent tests to scrutinise the progress of adaptation. In the ideal adapter framework, listener expectations are predicted to be rationally updated through integration with the incoming speech input and thus can theoretically be analysed on a trial-by-trial basis. The overall design of the studies reported here were motivated by our aim to understand this incremental belief-updating process which has not been closely studied in previous work. We thus address the limitations of previous work and in conjunction, make use of ideal observer models to validate baseline assumptions that accompany this kind of speech perception study – that listeners hold prior expectations or beliefs about cue distributions based on previously experienced speech input (here taken to mean native AE listeners' lifetime of experience with AE). Arriving at a definitive conclusion of what shape and form those beliefs take is beyond the scope of this study however we attempt to explore the

various proposals that have emerged from more than half a century of speech perception research.

A secondary aim was to begin to address possible concerns of ecological validity of prior work. While no speech stimuli is ever ideal, previous work on which the current study is based did have limitations in one or two aspects:the artificiality of the stimuli or the artificiality of the distributions. For e.g. (Clayards et al., 2008a) and (Kleinschmidt & Jaeger, 2016) made use of synthesised stimuli that were robotic or did not sound human-like. The second way that those studies were limited was that the exposure distributions of the linguistic categories had identical variances (see also Theodore & Monto, 2019) unlike what is found in production data where the variance of the voiceless categories are typically wider than that of the voiced category (Chodroff & Wilson, 2017). We take modest steps to begin to improve the ecological validity of this study while balancing the need for control through lab experiments by employing more natural sounding stimuli as well as by setting the variances of our exposure distributions to better reflect empirical data on production (see section x.xx. of SI).

2 Experiment 1: Listener's expectations prior to informative exposure

Experiment 1 investigates native (L1) US English listeners' categorization of word-initial stop voicing by an unfamiliar female L1 US English talker, prior to more informative exposure. Specifically, listeners heard isolated recordings from a /d/-/t/ continuum, and had to respond which word they heard (e.g., "din" or "tin"). The recordings varied in voice onset time (VOT), the primary phonetic cue to word-initial stop voicing in L1 US English, as well as correlated secondary cues (f0 and rhyme duration). Critically, exposure was relatively uninformative about the talker's use of the phonetic cues in that all phonetic realizations occurred equally often. The design of Experiment 1 serves two goals.

The first goal is methodological. We use Experiment 1 to test basic assumptions about the paradigm and stimuli we employ in the remainder of this study. We obtain estimates of the category boundary between /d/ and /t/ *for the specific stimuli used in Experiment 2*, as perceived by *the type of listeners we seek to recruit for Experiment 2*. We also test whether prolonged

testing across the phonetic continuum changes listeners' categorization behavior. Previous work has found that prolonged testing on uniform distributions can reduce the effects of previous exposure (Liu & Jaeger, 2018a; e.g., **mitterer2011?**), at least in listeners of the age group we recruit from (**scharenborg-janse2013?**). However, these studies employed only a small number of 5-7 perceptually highly ambiguous stimuli, each repeated many times. In Experiment 1, we employ a much larger set of stimuli that span the entire continuum from very clear /d/s to very clear /t/s, each presented only twice. If prolonged testing changes listeners' responses, this has to be taken into account in the design of Experiment 2.

The second purpose of Experiment 1 is to introduce and illustrate relevant theory. We compare different models of listeners' prior expectations against listeners' categorization responses in Experiment 1. The different models all aim to capture the implicit expectations of an L1 adult listener of US English might have about the mapping from acoustic cues to /d/ and /t/ based on previously experienced speech input. As we describe in more detail after the presentation of the experiment, the models differ, however, in whether these prior expectations take into account that talkers can differ in the way they realize /d/ and /t/. This ability to take into account talker differences even prior to more informative exposure is predicted—though through qualitatively different mechanisms, as we discuss below—both by normalization accounts (Cole, Linebaugh, Munson, & McMurray, 2010; McMurray & Jongman, 2011) and by accounts that attribute adaptive speech perception to changes in category representations (Bayesian ideal adaptor theory, Kleinschmidt & Jaeger, 2015; EARSHOT, Magnuson et al., 2020; episodic theory, Goldinger, 1998; exemplar theory, Johnson, 1997; Pierrehumbert, 2001). It is, however, unexpected under accounts that attribute adaptive speech perception solely to ad-hoc changes in decision-making. We did not expect that Experiment 1 yields a decisive conclusion with regard to this second goal, which is also addressed in Experiment 2. Rather, we use Experiment 1 as a presentationally convenient way to introduce some of the different models and provide readers with initial intuitions about what experiments of this type can and cannot achieve.

2.1   Methods

2.1.1   Participants

Participants were recruited over Amazon's Mechanical Turk platform, and paid $2.50 each (for a targeted remuneration of $6/hour). The experiment was only visible to Mechanical Turk participants who (1) had an IP address in the United States, (2) had an approval rating of 95% based on at least 50 previous assignments, and (3) had not previously participated in any experiment on stop voicing from our lab.

24 L1 US English listeners (female = 9; mean age = 36.2 years; SD age = 9.2 years) completed the experiment. To be eligible, participants had to confirm that they (1) spent at least the first 10 years of their life in the US speaking only English, (2) were in a quiet place, and (3) wore in-ear or over-the-ears headphones that cost at least $15.

2.1.2   Materials

We recorded multiple tokens of four minimal word pairs ("dill"/"till", "dim"/"tim", "din"/"tin", and "dip"/"tip") from a 23-year-old, female L1 US English talker with a mid-Western accent. These recordings were used to create four natural-sounding minimal pair VOT continua (dill-till, dip-tip, din-tin, and dip-tip) using a Praat script (Winn, 2020). The full procedure is described in the supplementary information (SI, **??**). The VOT continua ranged from -100ms VOT to +130ms VOT in 5ms steps. Experiment 1 employs 24 of these steps (-100, -50, -10, 5 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 110, 120, 130). VOT tokens in the lower and upper ends were distributed over larger increments because stimuli in those ranges were expected to elicit floor and ceiling effects, respectively.

We further set the F0 at vowel onset to follow the speaker's natural correlation which was estimated through a linear regression analysis of all the recorded speech tokens. We did this so that we could determine the approximate corresponding f0 values at each VOT value along the continua as predicted by this talker's VOT. The duration of the vowel was set to follow the natural trade-off relation with VOT reported in Allen and Miller (1999). This approach closely resembles that taken in Theodore and Monto (2019), and resulted in continuum steps that sound highly natural (unlike the robotic-sounding stimuli employed in Clayards et al., 2008a; Kleinschmidt & Jaeger, 2016). All stimuli are available as part of the OSF repository for this article.

196   In addition to the critical minimal pair continua we also recorded three words that did not

197 did not contain any stop consonant sounds ("flare", "share", and "rare"). These word recordings

198 were used as catch trials. Stimulus intensity was set to 70 dB sound pressure level for all

199 recordings.

2.1.3 Procedure

201 The code for the experiment is available as part of the OSF repository for this article. A live

202 version is available at (https://www.hlp.rochester.edu/FILLIN-FULL-URL). The first page of the

203 experiment informed participants of their rights and the requirements for the experiment: that

204 they had to be native listeners of English, wear headphones for the entire duration of the

205 experiment, and be in a quiet room without distractions. Participants had to pass a headphone

206 test, and were asked to keep the volume unchanged throughout the experiment. Participants could

207 only advance to the start of the experiment by acknowledging each requirement and consenting to

208 the guidelines of the Research Subjects Review Board of the University of Rochester.

209   On the next page, participants were informed about the task for the remainder of the

210 experiment. They were informed that they would heard a female talker speak a single word on

211 each trial, and had to select which word they heard. Participants were instructed to listen

212 carefully and answer as quickly and as accurately as possible. They were also alerted to the fact

213 that the recordings were subtly different and therefore may sound repetitive. This was done to

214 encourage their full attention.

215   Each trial started with a dark-shaded green fixation dot being displayed. At 500ms from

216 trial onset, two minimal pair words appeared on the screen, as shown in Figure 1. At 1000ms

217 from trial onset, the fixation dot would turn bright green and an audio recording from the

218 matching minimal pair continuum started playing. Participants were required to click on the

219 word they heard. For each participant, /d/-initial words were either always displayed on the left

220 side or always displayed on the right side. Across participants, this ordering was counter-balanced.

221 After participants clicked on the word, the next trial began.

222   Participants heard 192 target trials (four minimal pair continua, each with 24 VOT steps,

223 each heard twice). In addition, participants heard 12 catch trials. On catch trials, participant saw
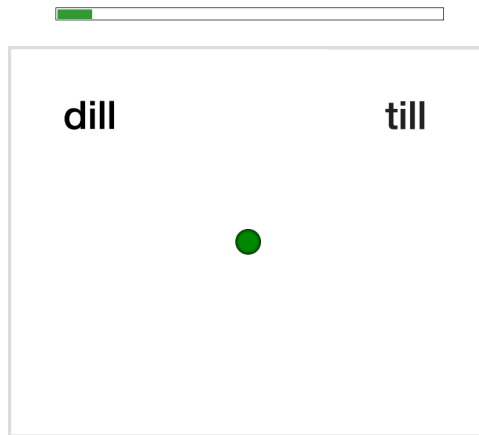
Figure 1. Example trial display. The words were displayed 500ms after trial onset and the audio recording of the word was played 1000ms after trial onset

two written catch stimuli on the screen (e.g., "flare" and "rare"), and heard one of them

(e.g. "rare"). Since these recordings were easily distinguishable, they served as a check on

participant attention throughout the experiment.

The order of trials was randomized for each participant with the only constraint that no

stimulus was repeated before each stimulus had been heard at least once. Catch trials were

distributed randomly throughout the experiment with the constraint that no more than two catch

trials would occur in a row. Participants were given the opportunity to take breaks after every 60

trials. Participants took an average of 12 minutes (SD = 4.8) to complete the 204 trials, after

which they answered a short survey about the experiment.

2.1.4 Exclusions

We excluded from analysis participants who committed more than 2 errors out of the 12 catch

trials ($<83\%$ accuracy, N = 3), participants with an average reaction time (RT) more than three

standard deviations from the mean of the by-participant means (N = 0), and participants who

reported not to have used headphones (N = 0) or not to be native (L1) speakers of US English (N

= 0). For the remaining participants, trials that were more than three SDs from the participant's

mean RT were excluded from analysis (1.6%). Finally, we excluded participants (N = 0) who had

less than 50% data remaining after these exclusions.

2.2   Behavioral results

We first present the behavioral analyses of participants' categorisation responses. Then we compare participants' responses to the predictions of different models fit on the distribution of stop voicing cues in a large database of L1 US English productions of word-initial /d/s and /t/s (Chodroff & Wilson, 2018).

2.2.1   Analysis approach

The goal of our behavioral analyses was to address three methodological questions that are of relevance to Experiment 2: (1) whether our stimuli resulted in 'reasonable' categorisation functions, (2) whether these functions differed between the four minimal pair items, and (3) whether participants' categorisation functions changed throughout the 192 test trials.

To address these questions, we fit a single Bayesian mixed-effects psychometric model to participants' categorization responses on critical trials (e.g., **prins2011?**). This model is essentially an extension of mixed-effects logistic regression that also takes into account attentional lapses. A failure to do so—while commonplace in research on speech perception (incl. our own work, but see Clayards, Tanenhaus, Aslin, & Jacobs, 2008b; Kleinschmidt & Jaeger, 2016)—can lead to biased estimates of categorization boundaries (e.g., **wichman-hill2001?**). The mixed-effects psychometric model describes the probability of "t"-responses as a weighted mixture of a lapsing-model and a perceptual model. The lapsing model is a mixed-effects logistic regression (Jaeger, 2008) that predicts participant responses that are made independent of the stimulus—for example, responses that result from attentional lapses. These responses are independent of the stimulus, and depend only on participants' response bias. The perceptual model is a mixed-effects logistic regression that predicts all other responses, and captures stimulus-dependent aspects of participants' responses. The relative weight of the two models is determined by the lapse rate, which is described by a third mixed-effects logistic regression.

The *lapsing model* only contained an intercept (the response bias in log-odds) and by-participant random intercepts. Similarly, the *model for the lapse rate* only had an intercept (the lapse rate) and by-participants random intercepts. No by-item random effects were included for the lapse rate nor lapsing model since these parts of the analysis—by definition—describe

stimulus-*in*dependent behavior. The *perceptual model* included an intercept and VOT, as well as the full random effect structure by participants and items (the four minimal pair continua), including random intercepts and random slopes by participant and minimal pair. We did not model the random effects of trial to reduce model complexity. This potentially makes our analysis of trials in the model anti-conservative. Finally, the models included the covariance between by-participant random effects across the three linear predictors for the lapsing model, lapse rate model, and perceptual model. This allows us to capture whether participants who lapse more often have, for example, different response biases or different sensitivity to VOT (after accounting for lapsing).

We fit the model using the package `brms` (Bürkner, 2017) in R (R Core Team, 2021a; RStudio Team, 2020). Following previous work from our lab (Hörberg & Jaeger, 2021; X. Xie et al., 2021), we used weakly regularizing priors to facilitate model convergence. For fixed effect parameters, we standardized continuous predictors (VOT) by dividing through twice their standard deviation (**gelman2008standardize?**), and used Student priors centered around zero with a scale of 2.5 units (following **gelman2008weakly?**) and 3 degrees of freedom. For random effect standard deviations, we used a Cauchy prior with location 0 and scale 2, and for random effect correlations, we used an uninformative LKJ-Correlation prior with its only parameter set to 1, describing a uniform prior over correlation matrices (**Lewandowski2009?**). Four chains with 2000 warm-up samples and 2000 posterior samples each were fit. No divergent transitions after warm-up were observed, and all $\hat{R}$ were close to 1.

2.2.2   Expectations

Based on previous experiments, we expected a strong positive effect of VOT, with increasing proportions of "t"-responses for increasing VOTs. We did not have clear expectations for the effect of trial other than that responses should become more uniformed (i.e move towards 50-50 "d"/"t"-bias or 0-log-odds) as the experiment progressed (Liu & Jaeger, 2018b) due to the un-informativeness of the stimuli. Previous studies with similar paradigms have typically found lapse rates of 0-10% ($<$ -2.2 log-odds, e.g., Clayards et al., 2008a; Kleinschmidt & Jaeger, 2016).

The lapse rate was estimated to be on the slightly larger side, but within the expected
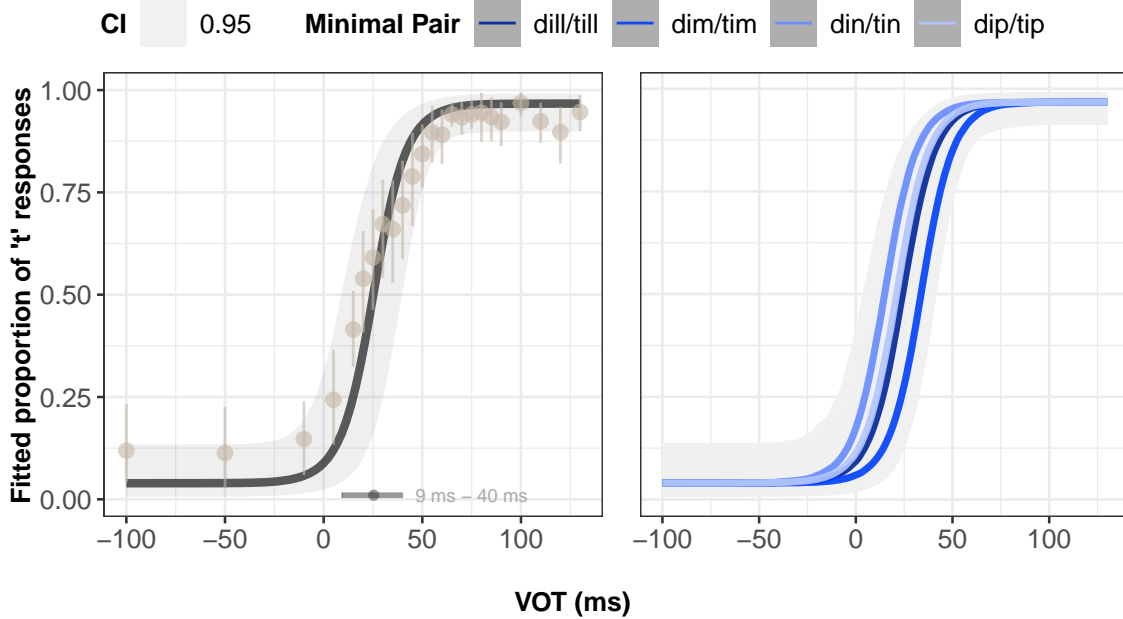
Figure 2. Categorisation functions and points of subjective equality (PSE) derived from the Bayesian mixed-effects psychometric model fit to listeners' responses in Experiment 1. The categorization functions include lapse rates and biases. The PSEs correct for lapse rates and lapse biases (i.e., they are the PSEs of the perceptual component of the psychometric model).[1] **Panel A:** Effects of VOT, lapse rate, and lapse bias, while marginalizing over trial effects as well as all random effects. Vertical point ranges represent the mean proportion and 95% bootstrapped CIs of participants' "t"-responses at each VOT step. Horizontal point ranges denote the mean and 95% quantile interval of the points of subjective equality (PSE), derived from the 8000 posterior samples of the population parameters. **Panel B:** The same but showing the fitted categorization functions for each of the four minimal pair continua. Participants' responses are omitted to avoid clutter. **Panel C:** Joint effects of VOT and trial as well as lapse rate and bias, while marginalizing over random effects.

297 range (7.5 %, 95%-CI: 2.2 to 21.2%; Bayes factor: 1,599 90%-CI : -3.54 to -1.53). Maximum a

298 posteriori (MAP) estimates of by-participant lapse rates ranged from XX . Very high lapse rates

299 were estimated for four of the participants with one in particular whose CI indicated exceptionally

300 high uncertainty. These lapse rates might reflect data quality issues with Mechanical Turk that

301 started to emerge over recent years (see **REFS?**; and, specifically for experiments on speech

302 perception, **cummings2023?**), and we return to this issue in Experiment 2.

303   The response bias were estimated to slightly favor "t"-responses (53.4 %, 95%-CI: 17.1 to

304 82.1%; Bayes factor: 1.52 90%-CI : -1.21 to 1.31), as also visible in Figure 2 (left). Unsurprisingly,

305 the psychometric model suggests high uncertainty about the participant-specific response biases,

306 as it is difficult to reliably estimate participant-specific biases while also accounting for trial and

VOT effects (range of by-participant MAP estimates: XX). For all but four participants, the 95%
CI includes the hypothesis that responses were unbiased. Of the remaining four participants,
three were biased towards "t"-responses and one was biased toward "d"-responses.

There was no convincing evidence of a main effect of trial ($\hat{\beta}$ = -0.2 95%-CI: -0.6 to 0.4;
Bayes factor: 2.71 90%-CI : -0.57 to 0.26). Given the slight overall bias towards "t"-responses, the
direction of this effect indicates that participants converged towards a 50/50 bias as the test
phase proceeded. This is also evident in Figure 2 (right). In contrast, there was clear evidence for
a positive main effect of VOT on the proportion of "t"-responses ($\hat{\beta}$ = 12.6 95%-CI: 9.8 to 15.5;
Bayes factor: Inf 90%-CI : 10.27 to 15.04). The effect of VOT was consistent across all minimal
pair words as evident from the slopes of the fitted lines by minimal pair 2 (left). MAP estimates
of by minimal pair slopes ranged from . The by minimal-pair intercepts were more varied (MAP
estimates:) with one of the pairs, dim/tim having a slightly lower intercept resulting in fewer
't'-responses on average. In all, this justifies our assumptions that word pair would not have a
substantial effect on categorisation behaviour. From the parameter estimates of the overall fit we
obtained the category boundary from the point of subjective equality (PSE) (25ms) which we use
for the design of Experiment 2.

Finally to accomplish the first goal of experiment 1, we look at the interaction between
VOT and trial. There was weak evidence that the effect of VOT decreased across trials ($\hat{\beta}$ = -0.6
95%-CI: -2.6 to 1.4; Bayes factor: 2.76 90%-CI : -2.27 to 1.05). The direction of this
change—towards more shallow VOT slopes as the experiment progressed—makes sense since the
test stimuli were not informative about the talker's pronunciation. Similar changes throughout
prolonged testing have been reported in previous work. (Liu & Jaeger, 2018a, 2019; **REFS?**).

Overall, there was little evidence that participants substantially changed their
categorisation behaviour as the experiment progressed. Still, to err on the cautious side,
Experiment 2 employs shorter test phases.

2.3   Comparisons to model of adaptive speech perception

We now turn to final aim of experiment 1 which is to make use of computational models to begin
to understand the implicit expectations that listeners hold when perceiving input that is

335 uninformative of a talker's cue-to-category-mappings.

336   Speakers' productions can act as a proxy for listeners' implicit knowledge of the

337 distributional patterns of cues. This production-perception relationship within a phonological

338 system was observed in early work by (Abramson & Lisker, 1973) who found that production

339 statistics of talkers along VOT aligned well with data from listeners who had categorised a

340 separate set of synthesised VOT stimuli. This allows for the use of analytic models as tools for

341 predicting categorisation behaviour from speech production data (Nearey & Hogan, 1986).

342   We apply this principle when fitting ideal observer (IO) models by linking the distributional

343 patterns of talker productions to the categorisation behaviour of listeners. All models were

344 trained on cue measurements extracted from an annotated database of 92 L1 US-English talkers'

345 productions (Chodroff & Wilson, 2017) of word initial /d/ and /t/. By using IOs trained solely

346 on production data to predict behaviour we avoid additional computational degrees of freedom

347 and limit the risk of overfitting the model to the data. We compare listener categorisation

348 behaviour against the predictions of five IO models which reflect different assumptions about

349 perceptual processes and the normalization (or lack thereof) of input. Beginning from the

350 minimal assumption of using raw VOT cues to predict categorisation, each successive model

351 builds up in complexity by cumulatively adding assumption over the previous model (Figure 3).

352   The first point that stands out from the visual comparisons is that models that incorporate

353 perceptual noise fit the perceptual data better than those that do not. This itself indicates that

354 perception of acoustic stimuli is not entirely faithful to the bottom-up signal but is inferred

355 through a combination of what listeners actually perceived and their existing knowledge of the

356 underlying linguistic category (Kronrod, Coppess, & Feldman, 2016). The second pattern is that

357 models trained with multiple cues are better fits overall than models trained on a single cue. This

358 trend is expected given the literature on

```
359  ## Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if `.name_re
360  ## i Using compatibility `.name_repair`.
361  ## i The deprecated feature was likely used in the MVBeliefUpdatr package.
362  ##    Please report the issue to the authors.
363  ## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
364  ## i Please use `linewidth` instead.
```
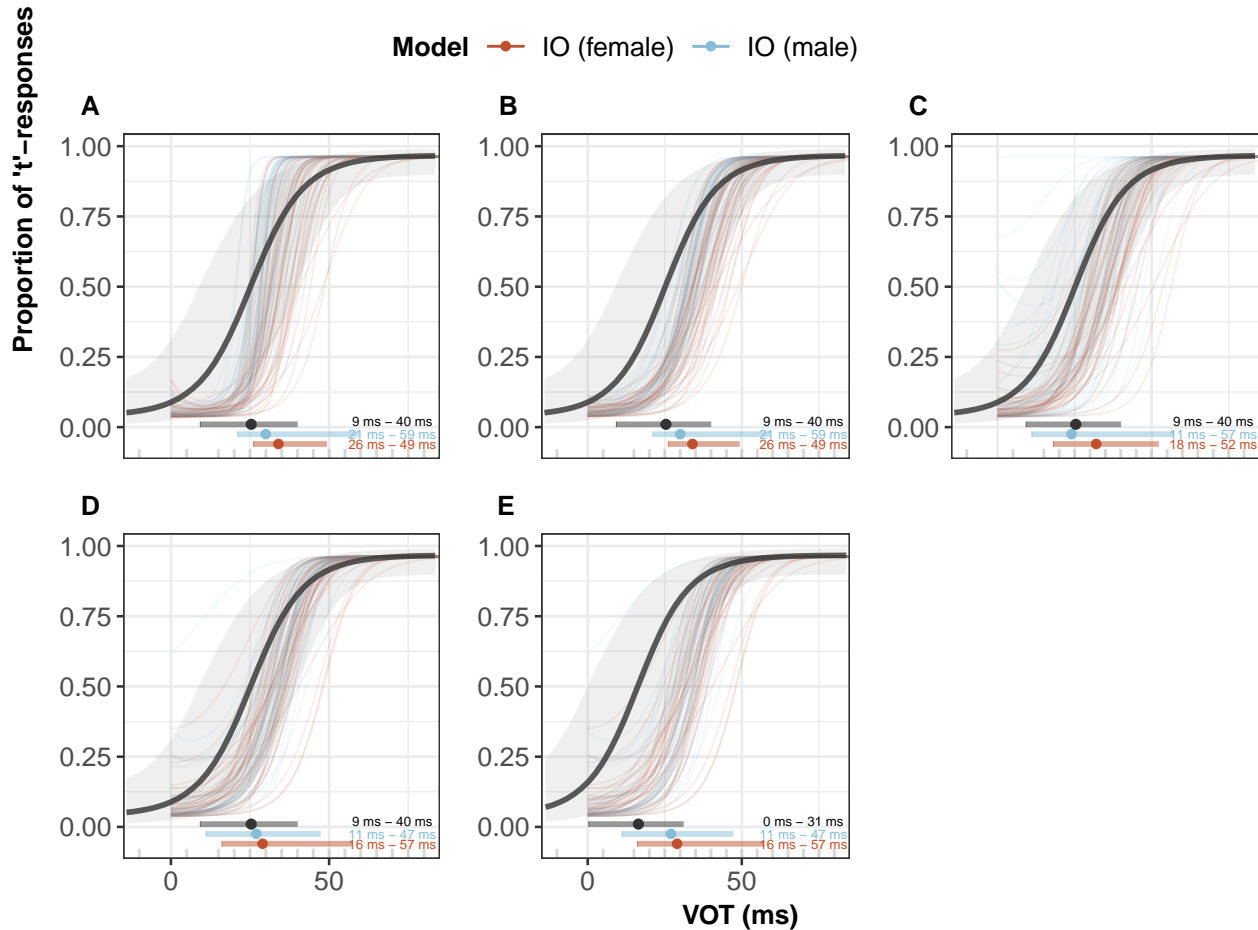

Figure 3. Comparing predicted vs. observed categorization functions for Experiment 1. The black line and interval show the psychometric fit and 95% CI for Experiment 1 marginalizing over all random effects. Each thin line shows the prediction of a single talker-specific ideal observers derived from a database of word-initial stop productions (data: Chodroff & Wilson, 2017; data preparation & model code: X. Xie, Jaeger, & Kurumada, 2022). The lapse rate and response bias for the ideal observers was set to match the MAP estimates of the psychometric model. For ease of comparisons, horizontal point ranges show the PSE and its 95% CI after discounting lapses.

365 **2.3.1   Comparing ideal observer accuracies**

366 Assess how well each of the four IOs fit human data.

367 ## Warning: Removed 2 rows containing missing values (`geom_point()`).

368 ## Removed 2 rows containing missing values (`geom_point()`).

Figure 4

3 EXPERIMENT 2: Listeners' adaptation to an unfamiliar talker

The aim of experiment 2 was to investigate the incremental changes in listener categorization when perceiving speech of an unfamiliar talker with cue-to-category mappings characterised by varying degrees of typ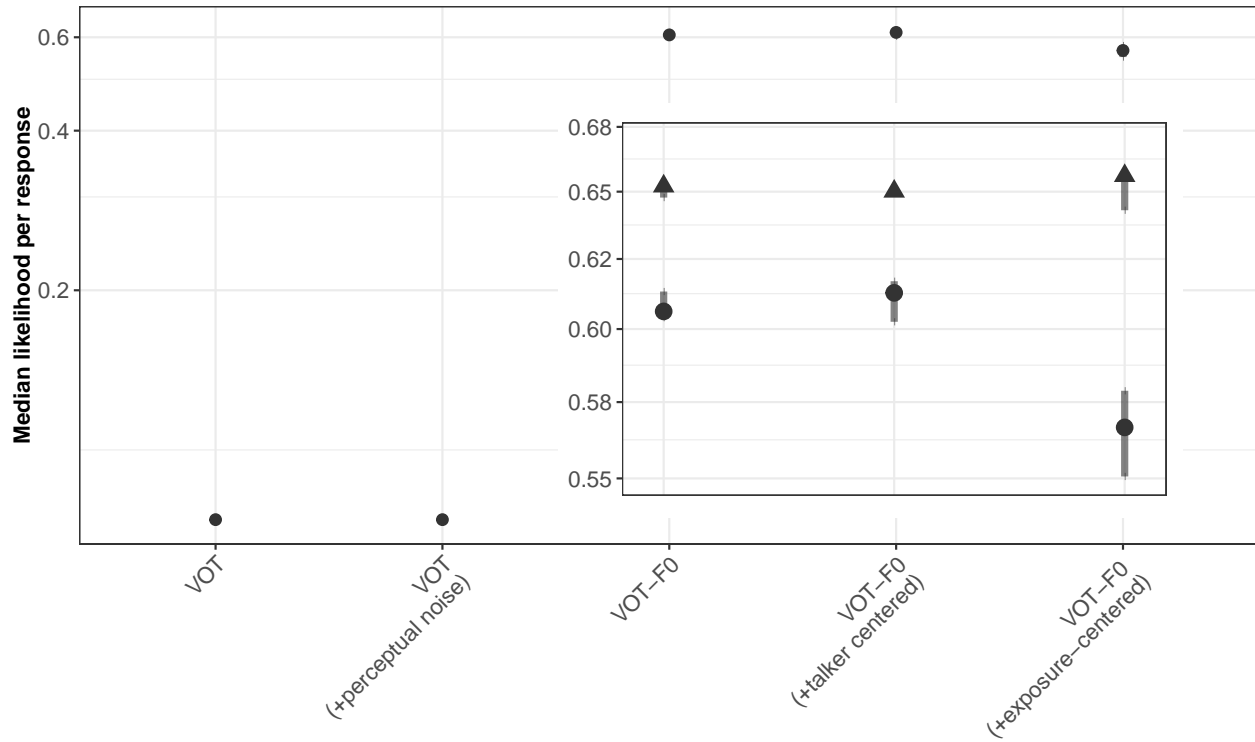icality of an L1-US English talker. Listeners performed a task similar to that of experiment 1, that is, they heard isolated words on a /d/ - /t/ continuum and were required to select the word they heard. Unlike experiment 1 where all listeners categorised stimuli on a single uninformative continuum, listeners in experiment 2 were divided into 3 groups with each group exposed to different VOT distributions that were informative of the talker's realisations of /d/ and /t/.

We approximated a "typical" talker through the combined parameters estimated from the perceptual responses in experiment 1 and a database of L1-US English /d/ and /t/ productions (**Xie?**). From this estimated baseline distribution (+0ms), we shifted the distribution by +10ms, and by +40ms, yielding three exposure talker conditions. To investigate the state of listener expectations as they move from having no information about how a new talker realises /d/s and /t/s to progressively more information about the talker's pronunciations we implement identical test blocks (i.e. test stimuli in identical locations) within and across conditions before, during, and after informative exposure. Under Bayesian ideal adaptor inferential processes, listeners' weighting of their prior beliefs about the category means and variances will determine the speed at which adaptation occurs. Motivated by prior work in supervised and unsupervised learning within lab contexts that repeatedly show adaptation to be a rapid process Kleinschmidt & Jaeger (2012) we made the decision to test our participants early on in the experiment.

Previous studies were not designed to investigate incremental adaptation in this manner as they lacked designated test blocks; listeners' categorisation functions were instead estimated over portions of the exposure trials which ignores the fact that not all participants had been exposed to the exact same items at the trial cut-off point (although that would have been the case by the end of the experiment). With our novel design we gain better resolution at every testing point, since each participant would have heard the same number of VOT items at the beginning of a

given test block. The other advantage is that identical test blocks across conditions standardises the assessment of behavioural changes between groups yielding more accurate comparisons. We specifically included a pre-exposure test block with a similar aim to experiment 1 – in order to capture the implicit expectations of listeners about the cue-to-category mappings of US English /d/ and /t/. We later compare this block with the behavioural results of experiment 1.

Previous studies found that listeners shift their categorization behaviour towards the category boundary implied by the exposure distribution but that adaptive shifts were incomplete, the further the exposure talker's distribution from a typical talker. We therefore expected to see differences in categorizations between the +10ms and +40ms conditions such that listeners in the +40ms condition would shift more than those in the +10ms but to have an average categorization function located to the left of the categorisation function that fully converges on the statistics of the exposure distribution. (Kleinschmidt & Jaeger, 2016). Nonetheless if adaptative speech perception involves rational updating we expect to find that the different shift conditions would induce changes in categorizations that are proportional to the distance between the shifts (i.e. +40ms being three times that of +10ms).

Another notable innovation we bring to this study in conjunction with the use of qualitatively more human-sounding stimuli (as described in section 2.X), relates to the parameters of the exposure distributions. Prior studies of this type simulate the voiced-voiceless distributions by exposing listeners to category distributions that are symmetrical and equivalent between categories. It is however, unlikely that listeners encounter this in real life as evidenced from production data (**chodroff?**). By generating distributions that are closer in form to that of real data we hope to improve the ecological validity of the results.

3.1 Methods

3.1.1 Participants

Participants were recruited over the Prolific platform and experiment data (but not participant profile data) were collected, stored, and via proliferate ((**schuster?**)). They were paid $8.00 each (for a targeted remuneration of $9.60/hour). The experiment was visible to participants following a selection of Prolific's available pre-screening criteria. Participants had to (1) have US

nationality, (2) report to only know English, and (3) had not previously participated in any experiment from our lab on Prolific.

126 L1 US English listeners (male = 60, female = 59, NA = 3; mean age = 38 years; SD age = 12 years) completed the experiment. Due to data transfer errors 4 participants' data were not stored and therefore not included in this analysis. To be eligible, participants had to confirm that they (1) spent at least the first 10 years of their life in the US speaking only English, (2) were in a quiet place and free from distractions, and (3) wore in-ear or over-the-ears headphones that cost at least $15.

3.1.2   Materials

A subset of the materials described in experiment 1 were used, in particular three continua of the minimal pairs, dill-till, din-tin, and dip-tip. The dim-tim continuum was omitted in order to keep the pairs as distinguishable as possible.

We employed a multi-block exposure-test design 5 which enabled the assessment of listener perception before informative exposure as well as incrementally at intervals during informative exposure (after every 48 exposure trials). To have a comparable test between blocks and across conditions, test blocks were made up of a uniform distribution of 12 VOT stimuli (-5, 5, 15, 25, 30, 35, 40, 45, 50, 55, 65, 70), identical across test blocks and between conditions. Each of the test tokens were presented once at random. The test blocks were kept short to avoid cancelling out any distributional learning effects after each exposure. After the final exposure block we tripled the number of test blocks to increase the statistical power to detect exposure induced changes.

The conditions were created by first obtaining the baseline distribution (+0ms shift) and then shifting that distribution by +10ms and by +40ms to create the remaining two conditions.

The +0ms shift condition was estimated from the fitted point of subjective equality (PSE) from experiment 1. The PSE corresponds to the VOT measurement that was perceived as the most ambiguous by participants in experiment 1 (i.e. the stimulus that elicited equal probability of being categorised as /d/ or /t/) thus marking the categorical boundary. The PSE is where the likelihoods of both categories intersect and have equal density (we assumed Gaussian distributions and equal prior probability for each category) [SOMETHING HERE ABOUT GAUSSIANS

BEING A CONVENIENT ASSUMPTION?]. To limit the infinite combinations of likelihoods
that could intersect at this value, we set the variances of the /d/ and /t/ categories based on
parameter estimates (X. Xie et al. (2022)) obtained from the production database of Chodroff
and Wilson (2017). To each variance value we added an 80ms noise variance following
((**kronrod?**)) to account for variability in perception due to perceptual noise since these
likelihoods were estimated from perceptual data. We took an additional degree of freedom of
setting the *distance between the means* of the categories at 46ms; this too was based on the
population parameter estimates from the production database. The means of both categories
were then obtained through a grid-search process to find the likelihood distributions that crossed
at 25ms VOT (see XX of SI for details on this procedure).

The distributional make up was determined through a process of sampling tokens from a
discrete normal distribution (available through the `extraDistr` package in R). [EXPLAIN WHAT
DISCRETE NORMAL SAMPLING GIVES] discretised normal distributions are approximation…

For each exposure block 8 VOT tokens of each minimal pair item were sampled from
discrete normal distributions of each category of the +0ms condition, giving 24 /d/ and 24 /t/
(48 critical trials) per block. Additionally, each exposure block contained 2 instances of 3 catch
items, giving 6 catch trials per block. The sampled VOT tokens were increased by a margin of
+10ms and +40 ms to create the remaining two conditions. Three variants of each condition list
were created so that exposure blocks followed a latin-square order.

Lastly, half of the exposure trials were randomly assigned as labelled trials. In labelled
trials, participants receive clear information of the word's category as both orthographic options
will always begin with the intended sound. For example if a trial was intended to be "dill" then
the two image options will either be "dill" and "dip" or "dill" and "din". Test trials were always
unlabelled.

3.1.3 Procedure

The code for the experiment is available as part of the OSF repository for this article. A live
version is available at (https://www.hlp.rochester.edu/FILLIN-FULL-URL). The first page of the
experiment informed participants of their rights and the requirements for the experiment: that
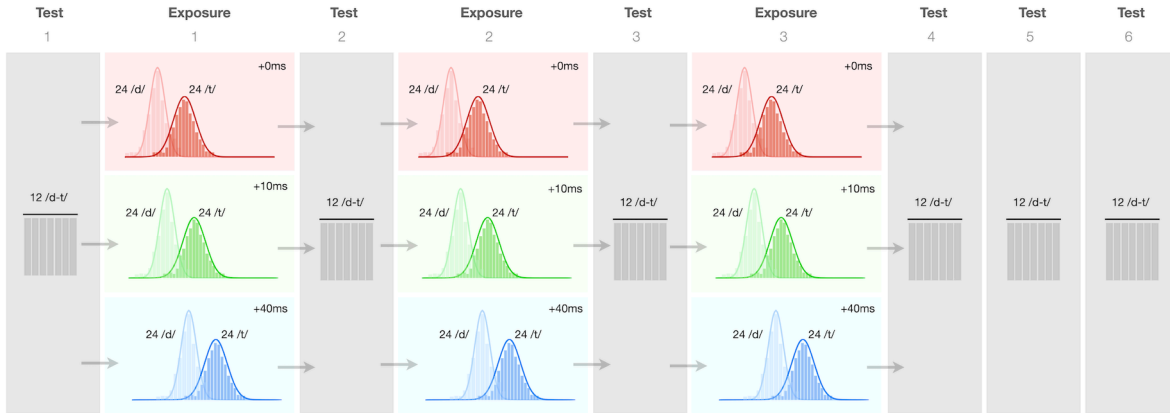
Figure 5. Experiment 2 multi-block design. Test blocks in grey comprised identical stimuli within and between conditions

481 they had to be native listeners of English, wear headphones for the entire duration of the

482 experiment, and be in a quiet room without distractions. Participants had to pass a headphone

483 test, and were asked to keep the volume unchanged throughout the experiment. Participants could

484 only advance to the start of the experiment by acknowledging each requirement and consenting to

485 the guidelines of the Research Subjects Review Board of the University of Rochester.

486 On the next page, participants were informed about the task for the remainder of the

487 experiment. They were informed that they would hear a female talker speak a single word on

488 each trial, and had to select which word they heard. They were also informed that they needed to

489 click a green button that would be displayed during each trial when it "lights up" in order to hear

490 the recording of the speaker saying the word. Participants were instructed to listen carefully and

491 answer as quickly and as accurately as possible. They were also alerted to the fact that the

492 recordings were subtly different and therefore may sound repetitive. This was done to encourage

493 their full attention.

494 The trials were presented in the same way as in experiment 1 except that the audio

495 playback was controlled by the participant. This additional step was implemented to increase

496 participant attention to the stimuli. The placement of the image presentations were

497 counter-balanced across participants.

498 Participants underwent 234 trials which included 6 catch trials in each exposure block (18

499 in total). Participants were given the opportunity to take breaks after every 60 trials during

Labelling /d/ labeled   /d/ unlabeled   /t/ labeled   /t/ unlabeled
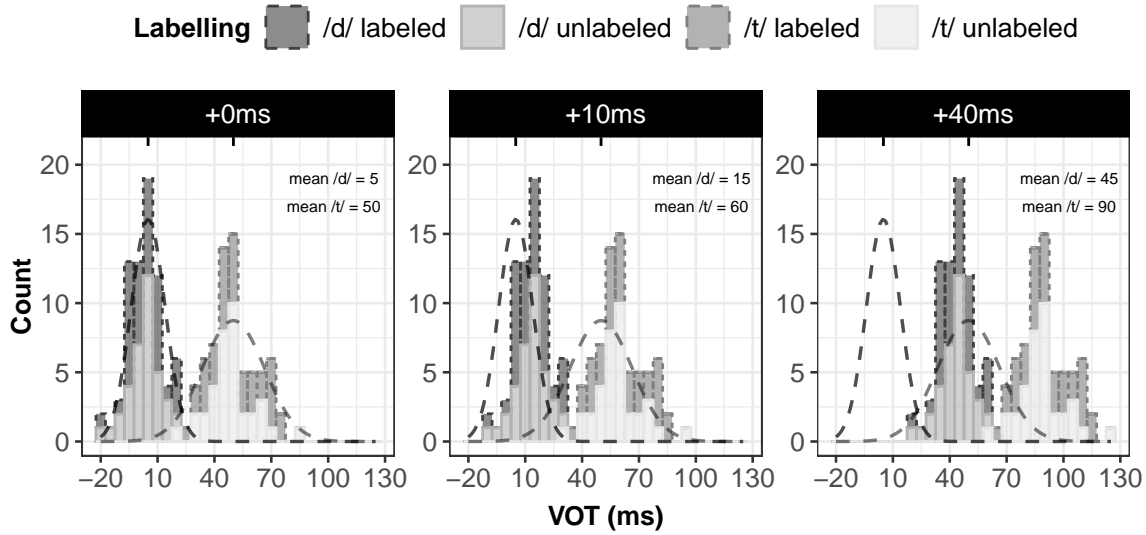
Figure 6

exposure blocks. Participants took an average of 17 minutes (SD = 9) to complete the 234 trials, after which they answered a short survey about the experiment.

```
## # A tibble: 5 x 3
## # Groups:   Exclude_participant.due_to_VOT_slope [2]
##    Exclude_participant.due_to_VOT_slope Condition.Exposure `n()`
##    <lgl>                                <chr>              <int>
## 1 FALSE                                 Shift0                41
## 2 FALSE                                 Shift10               40
## 3 FALSE                                 Shift40               39
## 4 TRUE                                  Shift0                 1
## 5 TRUE                                  Shift10                1
```

3.1.4 Exclusions

We excluded from analysis participants who committed more than 3 errors out of the 18 catch trials (<84% accuracy, N = 1), participants who committed more than 4 errors out of the 72 catch trials (<94% accuracy, N = 0), participants with an average reaction time (RT) more than three standard deviations from the mean of the by-participant means (N = 0), and participants who reported not to have used headphones (N = 0) or not to be native (L1) speakers of US

English (N = 0).

In addition, participants' categorization during the early phase of the experiment were scrutinised for their slope orientation and their proportion of "t"-responses at the least ambiguous locations of the VOT continuum. The early phase of the experiment was defined as the first 36 trials and the least ambiguous locations were defined as -20ms from the empirical mean of the /d/ category and +20ms from the empirical mean of the /t/ category. These means were taken from the production data estimates by X. Xie et al. (2022). For the remaining participants, trials that were more than three SDs from the participant's mean RT were excluded from analysis (0.012%). Finally, we excluded participants (N = 0) who had less than 50% data remaining after these exclusions.

All data and code for this article can be downloaded fromhttps://osf.io/q7gjp/. This article is written in R markdown, allowing readers to replicate our analyses with the press of a button using freely available software (R, R Core Team, 2021a; RStudio Team, 2020), while changing any of the parameters of our models. Readers can revisit any of the assumptions we make—for example, by substituting alternative models of linguistic representations. The supplementary information (SI, §1) lists the software/libraries required to compile this document. Beyond our immediate goals here, we hope that this can be helpful to researchers who are interested in developing more informative experimental designs, and to facilitate the interpretation of existing results (see also Tan, Xie, & Jaeger, 2021).

4   General discussion

Fig. XX summarizes participants' categorization functions across the different test blocks. To analyse the incremental effects of exposure condition on the proportion of /t/ responses at test, we fitted a Bayesian mixed-effects psychometric model with lapse rate (cf. Wichmann & Hill, 2001). The perceptual model contained exposure condition (sliding difference coded, comparing the +10ms against the +0ms shift condition, and the +40ms against the +10ms shift condition), test block (sliding difference coded from the first to last test block), VOT (Gelman scaled), and their full factorial interaction. We also included the full random effect structure by participant and item. The lapse rate and response bias (.5 for both /d/ and /t/) were assumed to be constant across blocks and exposure condition. We used the same weakly regularizing priors as in Xie, Liu, and Jaeger (2021). Condition and test blocks were successive-difference coded. There was a main effect of VOT; participants were more likely to give voiceless responses as VOT increased. Condition had a main effect on responses such that with larger shifts, participants on average responded with fewer /t/s. Additionally, the difference in average /t/ responses between the +40 and +10 conditions (-2.4 reduction in log-odds) was larger than the difference between the +10 and +0 conditions (-1.05 in log-odds). Qualitatively, the results indicate listeners adjust their expectations to align with the statistics of the exposure talker, consonant with previous findings of studies employing this paradigm (e.g., Clayards et al.; K&J16).

4.1 Methodological advances that can move the field forward

An example of a subsection.

5   References

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in spanish word-initial
    stops. *Journal of Phonetics*, *1*(1), 1–8.

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on
    the temporal characteristics of monosyllabic words. *The Journal of the Acoustical
    Society of America*, *106*(4), 2031–2039.

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.*
    Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2020). *Magrittr: A forward-pipe operator for r.* Retrieved
    from https://CRAN.R-project.org/package=magrittr

Barth, M. (2022). *tinylabels: Lightweight variable labels.* Retrieved from
    https://cran.r-project.org/package=tinylabels

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

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods.*
    Retrieved from https://CRAN.R-project.org/package=Matrix

Bolker, B., & Robinson, D. (2022). *Broom.mixed: Tidying methods for mixed models.*
    Retrieved from https://CRAN.R-project.org/package=broom.mixed

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech.
    Cognition, *106*(2), 707–729.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
    Journal of Statistical Software, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
    The R Journal, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan.
    Journal of Statistical Software, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization:
    Covariation of stop consonant VOT in american english. *Journal of Phonetics*, *61*,

585   30–47.

586   Chodroff, E., & Wilson, C. (2018). Predictability of stop consonant phonetics across

587   talkers: Between-category and within-category dependencies among cues for place and

588   voice. *Linguistics Vanguard*, *4*. https://doi.org/10.1515/lingvan-2017-0047

589   Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented english.

590   *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

591   Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008b). Perception of

592   speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809.

593   https://doi.org/10.1016/j.cognition.2008.04.004

594   Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008a). Perception of

595   speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

596   https://doi.org/https://doi.org/10.1016/j.cognition.2008.04.004

597   Cole, J., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic

598   effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of*

599   *Phonetics*, *38*, 167–184. https://doi.org/10.1016/j.wocn.2009.08.004

600   Csárdi, G., & Chang, W. (2021). *Processx: Execute and control system processes.*

601   Retrieved from https://CRAN.R-project.org/package=processx

602   Daróczi, G., & Tsegelskyi, R. (2022). *Pander: An r 'pandoc' writer.* Retrieved from

603   https://CRAN.R-project.org/package=pander

604   Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of 'data.frame'.* Retrieved from

605   https://CRAN.R-project.org/package=data.table

606   Eddelbuettel, D., & Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief

607   Introduction to extitRcpp. *The American Statistician*, *72*(1), 28–36.

608   https://doi.org/10.1080/00031305.2017.1375990

609   Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal*

610   *of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

611   Frick, H., Chow, F., Kuhn, M., Mahoney, M., Silge, J., & Wickham, H. (2022). *Rsample:*

612   *General resampling infrastructure.* Retrieved from

613   https://CRAN.R-project.org/package=rsample

614   Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access.

Psychological Review, *105*(2), 251.

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

Hay, J., & Drager, K. (2010). *Stuffed toys and speech perception.*

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Henry, L., & Wickham, H. (2021). *Rlang: Functions for base types and core r and 'tidyverse' features.* Retrieved from https://CRAN.R-project.org/package=rlang

Henry, L., Wickham, H., & Chang, W. (2020). *Ggstance: Horizontal 'ggplot2' components.* Retrieved from https://CRAN.R-project.org/package=ggstance

Hörberg, T., & Jaeger, T. F. (2021). A rational model of incremental argument interpretation: The comprehension of swedish transitive clauses. *Frontiers in Psychology*, *12*, 674202.

Hugh-Jones, D. (2021). *Latexdiffr: Diff 'rmarkdown' files using the 'latexdiff' utility.* Retrieved from https://CRAN.R-project.org/package=latexdiffr

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Johnson, K. (1997). Speech perception without speaker NormalizationÖ an exemplar model. *Talker Variability in Speech Processing*, 145–165.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots.* Retrieved from https://CRAN.R-project.org/package=ggpubr

Kay, M. (2022a). *ggdist: Visualizations of distributions and uncertainty.* https://doi.org/10.5281/zenodo.3879620

Kay, M. (2022b). *tidybayes: Tidy data and geoms for Bayesian models.* https://doi.org/10.5281/zenodo.1308151

Kleinschmidt, D. F., & Jaeger, T. F. (2012). A continuum of phonetic adaptation:

Evaluating an incremental belief-updating model of recalibration and selective adaptation. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148. https://doi.org/https://doi.org/10.1037/a0038695

Kleinschmidt, D. F., & Jaeger, T. F. (2016). What do you expect from an unfamiliar talker? *CogSci.*

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review, 23*(6), 1681–1712. https://doi.org/https://doi.org/10.3758/s13423-016-1049-y

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Liao, Y. (2019). *Linguisticsdown: Easy linguistics document writing with r markdown.* Retrieved from https://CRAN.R-project.org/package=linguisticsdown

Liu, L., & Jaeger, T. F. (2018a). Inferring causes during speech perception. *Cognition, 174*, 55–70. https://doi.org/10.1016/j.cognition.2018.01.003

Liu, L., & Jaeger, T. F. (2018b). Inferring causes during speech perception. *Cognition, 174*, 55–70.

Liu, L., & Jaeger, T. F. (2019). Talker-specific pronunciation or speech error? Discounting (or not) atypical pronunciations during speech perception. *Journal of Experimental Psychology. Human Perception and Performance, 45*, 1562–1588. https://doi.org/10.1037/xhp0000693

Maechler, M. (2021). *Diptest: Hartigan's dip test statistic for unimodality - corrected.* Retrieved from https://CRAN.R-project.org/package=diptest

Magnuson, J. S., You, H., Luthra, S., Li, M., Nam, H., Escabi, M., et al.others. (2020). EARSHOT: A minimal neural network model of incremental human speech recognition. *Cognitive Science, 44*(4), e12823.

McCloy, D. R. (2016). *phonR: Tools for phoneticians and phonologists.*

McMurray, B., & Jongman, A. (2011). What information is necessary for speech

675   categorization? Harnessing variability in the speech signal by integrating cues

676   computed relative to expectations. *Psychological Review, 118*(2), 219.

677   Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames.* Retrieved from

678   https://CRAN.R-project.org/package=tibble

679   Nearey, T. M., & Hogan, J. T. (1986). Phonological contrast in experimental phonetics:

680   Relating distributions of production data to perceptual categorization curves.

681   *Experimental Phonology*, 141–161.

682   Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes.* Retrieved from

683   https://CRAN.R-project.org/package=RColorBrewer

684   Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive*

685   *Psychology, 47*(2), 204–238.

686   Ooms, J. (2021). *Magick: Advanced graphics and image-processing in r.* Retrieved from

687   https://CRAN.R-project.org/package=magick

688   Ooms, J. (2022). *Curl: A modern and flexible web client for r.* Retrieved from

689   https://CRAN.R-project.org/package=curl

690   Pedersen, T. L. (2022a). *Ggforce: Accelerating 'ggplot2'.* Retrieved from

691   https://CRAN.R-project.org/package=ggforce

692   Pedersen, T. L. (2022b). *Patchwork: The composer of plots.* Retrieved from

693   https://CRAN.R-project.org/package=patchwork

694   Pedersen, T. L., & Robinson, D. (2020). *Gganimate: A grammar of animated graphics.*

695   Retrieved from https://CRAN.R-project.org/package=gganimate

696   Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast.

697   In J. Bybee & P. Hopper (Eds.), *In Frequency and the Emergence of Linguistic*

698   *Structure* (pp. 137–157). John Benjamins.

699   R Core Team. (2021a). *R: A language and environment for statistical computing.* Vienna,

700   Austria: R Foundation for Statistical Computing. Retrieved from

701   https://www.R-project.org/

702   R Core Team. (2021b). *R: A language and environment for statistical computing.* Vienna,

703   Austria: R Foundation for Statistical Computing. Retrieved from

704   https://www.R-project.org/

RStudio Team. (2020). *RStudio: Integrated development environment for r.* Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny.* Chapman; Hall/CRC. Retrieved from https://plotly-r.com

Slowikowski, K. (2021). *Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'.* Retrieved from https://CRAN.R-project.org/package=ggrepel

Statisticat, & LLC. (2021). *LaplacesDemon: Complete environment for bayesian inference.* Bayesian-Inference.com. Retrieved from https: //web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software

Tan, M., Xie, X., & Jaeger, T. F. (2021). Using rational models to understand experiments on accent adaptation. *Frontiers in Psychology*, *12*, 1–19. https://doi.org/10.3389/fpsyg.2021.676271

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *128*(4), 2090–2099.

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992. https://doi.org/https://doi.org/10.3758/s13423-018-1551-5

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved rhat for assessing convergence of MCMC (with discussion). *Bayesian Analysis.*

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from https://www.stats.ox.ac.uk/pub/MASS4/

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019a). *Assertthat: Easy pre and post assertions.* Retrieved from https://CRAN.R-project.org/package=assertthat

Wickham, H. (2019b). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2020). *Modelr: Modelling functions that work with the pipe.* Retrieved from

735   https://CRAN.R-project.org/package=modelr

736   Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors).*

737   Retrieved from https://CRAN.R-project.org/package=forcats

738   Wickham, H. (2021b). *Tidyr: Tidy messy data.* Retrieved from

739   https://CRAN.R-project.org/package=tidyr

740   Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani,

741   H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

742   https://doi.org/10.21105/joss.01686

743   Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data*

744   *manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

745   Wickham, H., Hester, J., & Bryan, J. (2021). *Readr: Read rectangular text data.*

746   Retrieved from https://CRAN.R-project.org/package=readr

747   Wickham, H., & Seidel, D. (2022). *Scales: Scale functions for visualization.* Retrieved

748   from https://CRAN.R-project.org/package=scales

749   Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.*

750   Retrieved from https://CRAN.R-project.org/package=cowplot

751   Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and

752   flexible praat script. *The Journal of the Acoustical Society of America*, *147*(2),

753   852–866.

754   Xie, X., Jaeger, T. F., & Kurumada, C. (2022). *What we do (not) know about the*

755   *mechanisms underlying adaptive speech perception: A computational review.*

756   https://doi.org/10.17605/OSF.IO/Q7GJP

757   Xie, X., Liu, L., & Jaeger, T. F. (2021). Cross-talker generalization in the perception of

758   nonnative speech: A large-scale replication. *Journal of Experimental Psychology:*

759   *General.*

760   Xie, X., Weatherholtz, K., Bainton, L., Rowe, E., Burchill, Z., Liu, L., & Jaeger, T. F.

761   (2018). Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar

762   talker. *The Journal of the Acoustical Society of America*, *143*(4), 2013–2031.

763   Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida:

764   Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/

765   Xie, Y. (2021). *Knitr: A general-purpose package for dynamic report generation in r.*

766   Retrieved from https://yihui.org/knitr/

767   Xie, Y., & Allaire, J. (2022). *Tufte: Tufte's styles for r markdown documents.* Retrieved

768   from https://CRAN.R-project.org/package=tufte

769   Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.*

770   Retrieved from https://CRAN.R-project.org/package=kableExtra

Supplementary information

Both the main text and these supplementary information (SI) are derived from the same R

markdown document available via OSF. It is best viewed using Acrobat Reader. Some links and

animations might not work in other PDF viewers.

§1   Required software

The document was compiled using `knitr` (Y. Xie, 2021) in RStudio with R:

```
##                   _
## platform       x86_64-apple-darwin17.0
## arch          x86_64
## os            darwin17.0
## system        x86_64, darwin17.0
## status
## major         4
## minor         1.3
## year          2022
## month         03
## day           10
## svn rev       81868
## language      R
## version.string R version 4.1.3 (2022-03-10)
## nickname      One Push-Up
```

You will also need to download the IPA font SIL Doulos and a Latex environment like (e.g.,

MacTex or the R library `tinytex`).

We used the following R packages to create this document: R (Version 4.1.3; R Core Team,

2021b) and the R-packages *}broom* [@}R-broom], *assertthat* (Version 0.2.1; Wickham, 2019a),

brms (Version 2.18.0; Bürkner, 2017, 2018, 2021), *broom.mixed* (Version 0.2.9.4; Bolker &

Robinson, 2022), *cowplot* (Version 1.1.1; Wilke, 2020), *curl* (Version 4.3.3; Ooms, 2022), *data.table*

(Version 1.14.6; Dowle & Srinivasan, 2021), *diptest* (Version 0.76.0; Maechler, 2021), *dplyr* (Version 1.0.10; Wickham, François, Henry, & Müller, 2021), *forcats* (Version 0.5.2; Wickham, 2021a), *gganimate* (Version 1.0.8; Pedersen & Robinson, 2020), *ggdist* (Version 3.2.0; Kay, 2022a), *ggforce* (Version 0.4.1; Pedersen, 2022a), *ggplot2* (Version 3.4.0; Wickham, 2016), *ggpubr* (Version 0.5.0; Kassambara, 2020), *ggrepel* (Version 0.9.2; Slowikowski, 2021), *ggstance* (Version 0.3.6; Henry, Wickham, & Chang, 2020), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.41; Y. Xie, 2015), *LaplacesDemon* (Version 16.1.6; Statisticat & LLC., 2021), *latexdiffr* (Version 0.1.0; Hugh-Jones, 2021), *linguisticsdown* (Version 1.2.0; Liao, 2019), *lme4* (Version 1.1.31; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *lubridate* (Version 1.9.0; Grolemund & Wickham, 2011), *magick* (Version 2.7.3; Ooms, 2021), *magrittr* (Version 2.0.3; Bache & Wickham, 2020), *MASS* (Version 7.3.58.1; Venables & Ripley, 2002), *Matrix* (Version 1.5.1; Bates & Maechler, 2021), *modelr* (Version 0.1.10; Wickham, 2020), *pander* (Version 0.6.5; Daróczi & Tsegelskyi, 2022), *papaja* (Version 0.1.1.9,001; Aust & Barth, 2020), *patchwork* (Version 1.1.2; Pedersen, 2022b), *phonR* (Version 1.0.7; McCloy, 2016), *plotly* (Version 4.10.1; Sievert, 2020), *posterior* (Version 1.3.1; Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2021), *processx* (Version 3.8.0; Csárdi & Chang, 2021), *purrr* (Version 0.3.5; Henry & Wickham, 2020), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *readr* (Version 2.1.3; Wickham, Hester, & Bryan, 2021), *rlang* (Version 1.0.6; Henry & Wickham, 2021), *rsample* (Version 1.1.1; Frick et al., 2022), *scales* (Version 1.2.1; Wickham & Seidel, 2022), *stringr* (Version 1.4.1; Wickham, 2019b), *tibble* (Version 3.1.8; Müller & Wickham, 2021), *tidybayes* (Version 3.0.2; Kay, 2022b), *tidyr* (Version 1.2.1; Wickham, 2021b), *tidyverse* (Version 1.3.2; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), and *tufte* (Version 0.12; Y. Xie & Allaire, 2022). If opened in RStudio, the top of the R markdown document should alert you to any libraries you will need to download, if you have not already installed them. The full session information is provided at the end of this document.

§2 Overview

§2.1 Overview of data organisation

§3   Stimuli generation for perception experiments

§3.1   Recording of audio stimuli

§3.2   Annotation of audio stimuli

§3.3   Synthesis of audio stimuli

- acoustic plots

§4   Web-based experiment design procedure

§4.1   Experiment 1

§4.1.1   Making exposure conditions

§4.1.2   Exclusions analysis

§4.2   Experiment 2

§4.2.1   Making exposure conditions

§4.2.2   Exclusions analysis

- reaction time plots
- catch trial performance plots

    -labelled trial performance plots

§4.3   Ideal observer training

We train the IOs on cue distributions extracted from an annotated database of XX L1 US-English talkers' productions (Chodroff and Wilson (2017)) of word initial stops. We apply Bayes' theorem to derive the IOs' posterior probability of categorising the test stimuli as "t". This is defined as the product of the likelihood of the cue under the hypothesis that the talker produced "t", and the prior probability of that cue. By using IOs trained solely on production data to predict categorization behaviour we avoid additional computational degrees of freedom and limit the risk of overfitting the model to the data thus reducing bias.
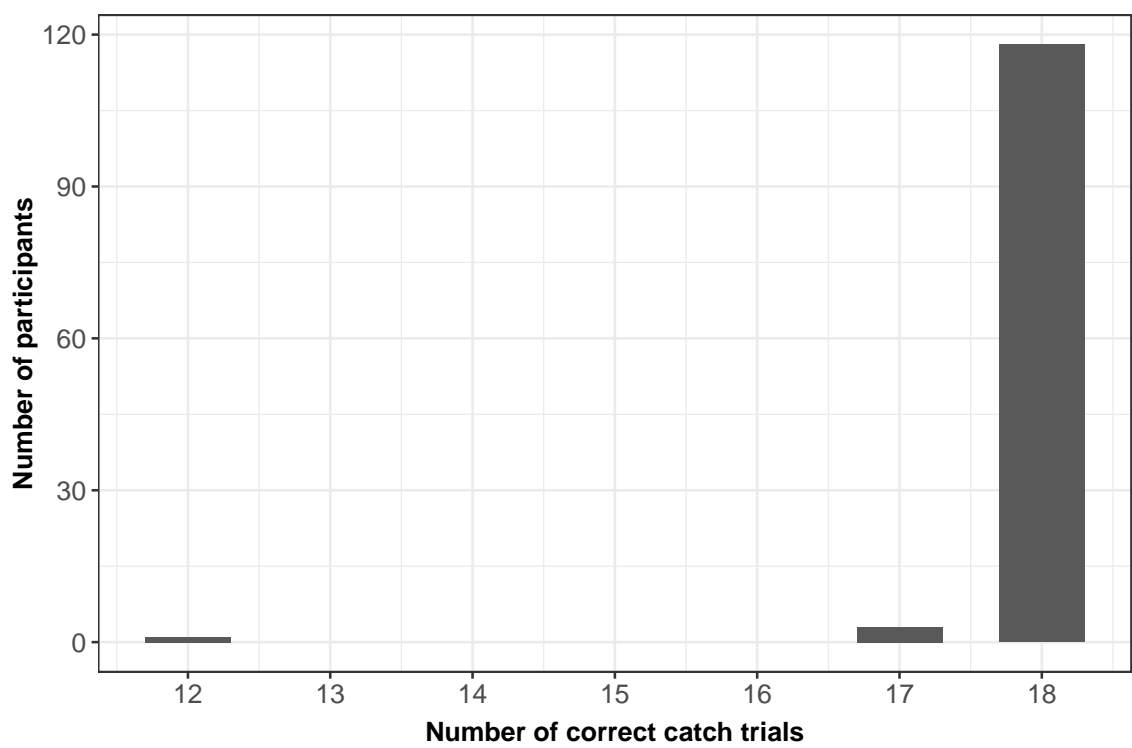
Figure 7
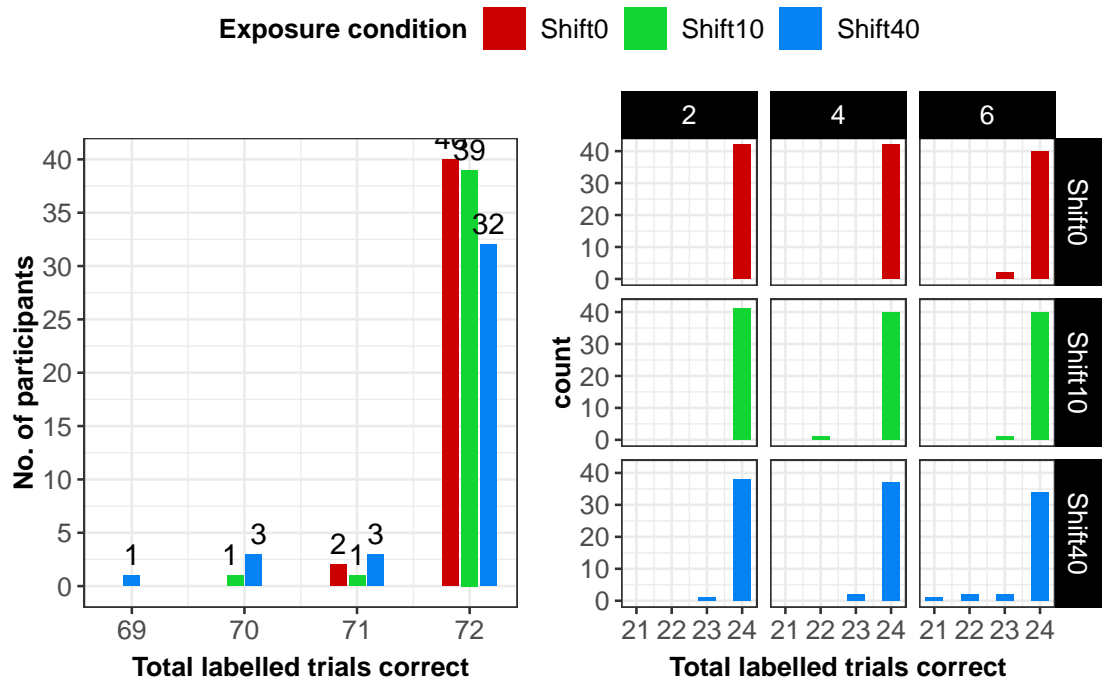
Figure 8
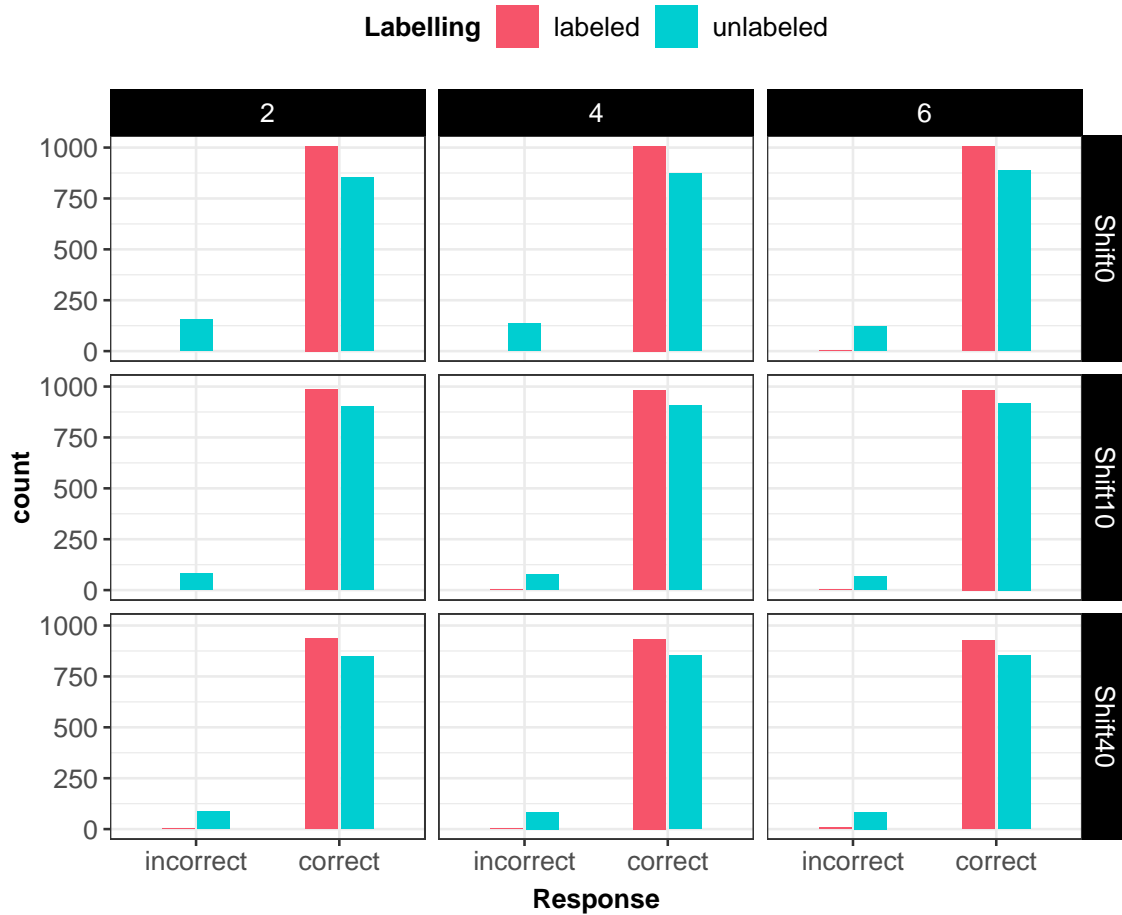
Figure 9

⁸⁴⁹   We filtered the database to /d/s and /t/s which gave 92 talkers (4x male and 4x female),

⁸⁵⁰ each with a minimum of 25 tokens. We then fit ideal observers to each talker under different

⁸⁵¹ hypotheses of distributional learning [and evaluated their respective goodness-of-fit to the human

⁸⁵² data]. In total we fit x IOs to represent the different hypotheses about listeners' implicit

⁸⁵³ knowledge – models grouped by sex, grouped by sex and Predictions of the IO were obtained

⁸⁵⁴ using talker-normalized category statistics for /d/ and /t/ from (X. Xie et al., 2022) based on

⁸⁵⁵ data from (**chodroff2017?**), perceptual noise estimates for VOT from (Kronrod et al., 2016), and

⁸⁵⁶ a lapse rate identical to the psychometric model estimate.

⁸⁵⁷ §5   Session Info

⁸⁵⁸ `## - Session info -----------------------------------------------------------------------------`
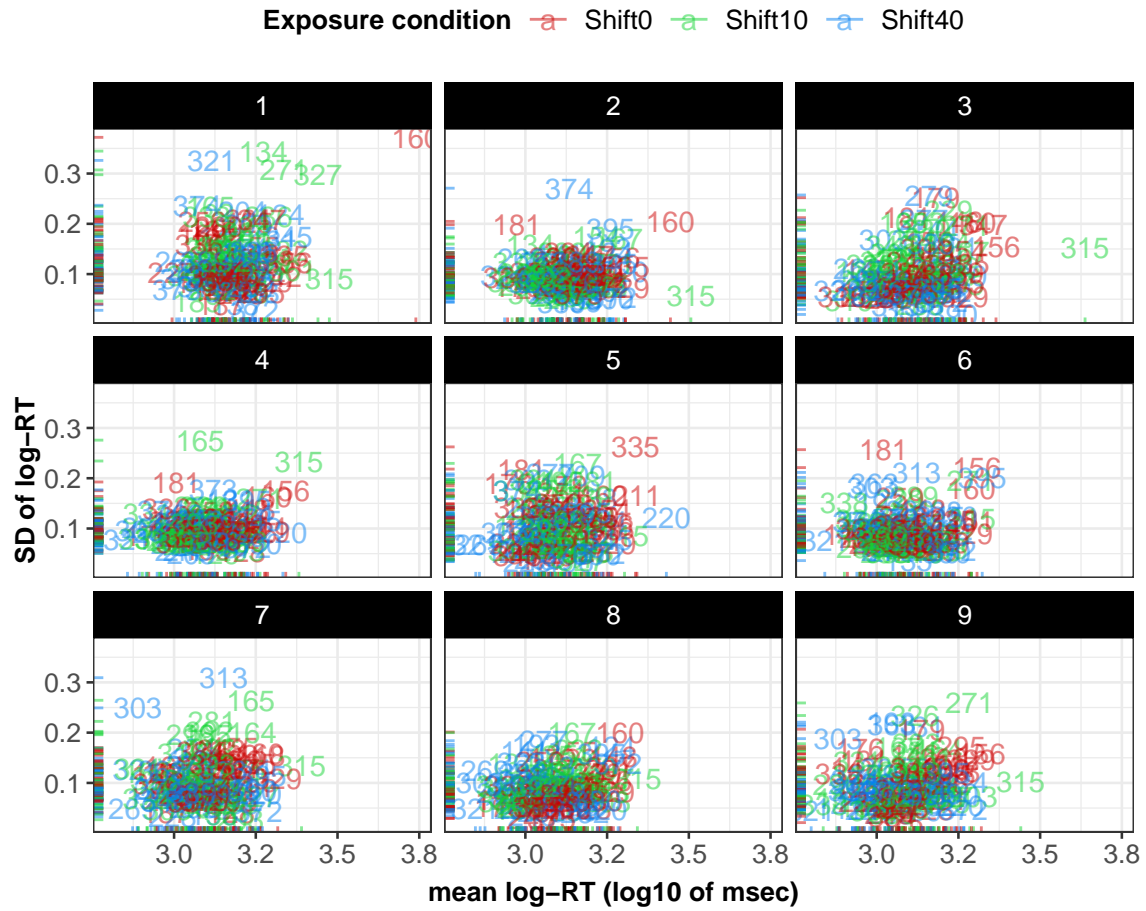
Figure 10
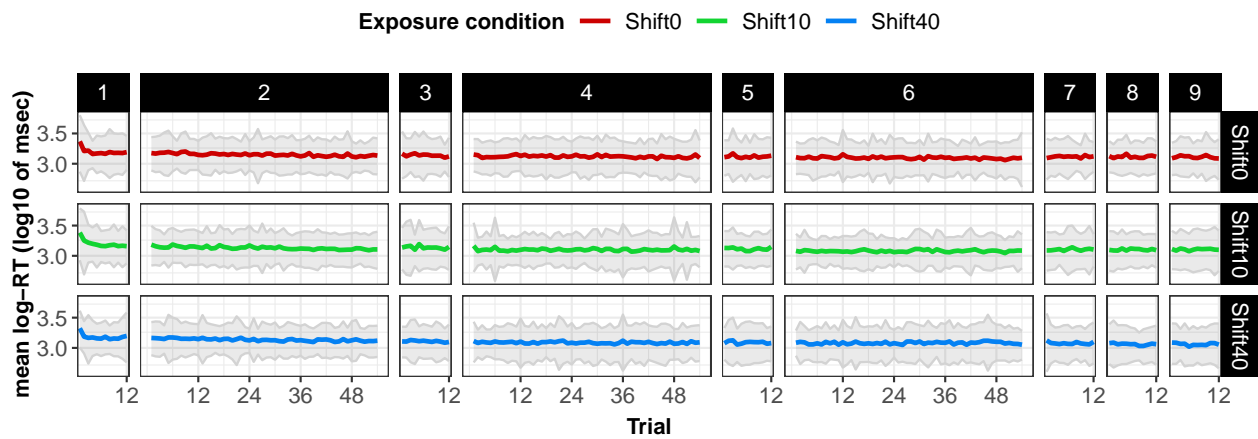
Figure 11

```
859  ##   setting  value
860  ##   version  R version 4.1.3 (2022-03-10)
861  ##   os       macOS Big Sur/Monterey 10.16
862  ##   system   x86_64, darwin17.0
863  ##   ui       X11
864  ##   language (EN)
865  ##   collate  en_US.UTF-8
866  ##   ctype    en_US.UTF-8
867  ##   tz       Europe/Stockholm
868  ##   date     2023-01-23
869  ##   pandoc   2.18 @ /Applications/RStudio.app/Contents/MacOS/quarto/bin/tools/ (via rmarkdown)
870  ##
871  ## - Packages ----------------------------------------------------------------------------
872  ##   package          * version   date (UTC) lib source
873  ##   abind              1.4-5     2016-07-21 [1] CRAN (R 4.1.0)
874  ##   arrayhelpers       1.1-0     2020-02-04 [1] CRAN (R 4.1.0)
875  ##   assertthat       * 0.2.1     2019-03-21 [1] CRAN (R 4.1.0)
876  ##   av                 0.8.2     2022-10-06 [1] CRAN (R 4.1.2)
877  ##   backports          1.4.1     2021-12-13 [1] CRAN (R 4.1.0)
878  ##   base64enc          0.1-3     2015-07-28 [1] CRAN (R 4.1.0)
879  ##   bayesplot          1.10.0    2022-11-16 [1] CRAN (R 4.1.2)
880  ##   bayestestR         0.13.0    2022-09-18 [1] CRAN (R 4.1.2)
881  ##   bit                4.0.5     2022-11-15 [1] CRAN (R 4.1.2)
882  ##   bit64              4.0.5     2020-08-30 [1] CRAN (R 4.1.0)
883  ##   bookdown           0.30      2022-11-09 [1] CRAN (R 4.1.2)
884  ##   boot               1.3-28.1  2022-11-22 [1] CRAN (R 4.1.2)
885  ##   bridgesampling     1.1-2     2021-04-16 [1] CRAN (R 4.1.0)
886  ##   brms             * 2.18.0    2022-09-19 [1] CRAN (R 4.1.2)
887  ##   Brobdingnag        1.2-9     2022-10-19 [1] CRAN (R 4.1.2)
888  ##   broom              1.0.1     2022-08-29 [1] CRAN (R 4.1.2)
```

```
889 ##   broom.mixed      0.2.9.4    2022-04-17 [1] CRAN (R 4.1.2)
890 ##   cachem           1.0.6      2021-08-19 [1] CRAN (R 4.1.0)
891 ##   callr            3.7.3      2022-11-02 [1] CRAN (R 4.1.2)
892 ##   car              3.1-1      2022-10-19 [1] CRAN (R 4.1.2)
893 ##   carData          3.0-5      2022-01-06 [1] CRAN (R 4.1.2)
894 ##   cellranger       1.1.0      2016-07-27 [1] CRAN (R 4.1.0)
895 ##   checkmate        2.1.0      2022-04-21 [1] CRAN (R 4.1.2)
896 ##   class            7.3-20     2022-01-16 [1] CRAN (R 4.1.3)
897 ##   classInt         0.4-8      2022-09-29 [1] CRAN (R 4.1.2)
898 ##   cli              3.4.1      2022-09-23 [1] CRAN (R 4.1.2)
899 ##   cluster          2.1.4      2022-08-22 [1] CRAN (R 4.1.2)
900 ##   coda             0.19-4     2020-09-30 [1] CRAN (R 4.1.0)
901 ##   codetools        0.2-18     2020-11-04 [1] CRAN (R 4.1.3)
902 ##   colorspace       2.0-3      2022-02-21 [1] CRAN (R 4.1.2)
903 ##   colourpicker     1.2.0      2022-10-28 [1] CRAN (R 4.1.2)
904 ##   cowplot        * 1.1.1      2020-12-30 [1] CRAN (R 4.1.0)
905 ##   crayon           1.5.2      2022-09-29 [1] CRAN (R 4.1.2)
906 ##   crosstalk        1.2.0      2021-11-04 [1] CRAN (R 4.1.0)
907 ##   curl           * 4.3.3      2022-10-06 [1] CRAN (R 4.1.2)
908 ##   data.table       1.14.6     2022-11-16 [1] CRAN (R 4.1.2)
909 ##   datawizard       0.6.4      2022-11-19 [1] CRAN (R 4.1.2)
910 ##   DBI              1.1.3      2022-06-18 [1] CRAN (R 4.1.2)
911 ##   dbplyr           2.2.1      2022-06-27 [1] CRAN (R 4.1.2)
912 ##   deldir           1.0-6      2021-10-23 [1] CRAN (R 4.1.0)
913 ##   devtools         2.4.5      2022-10-11 [1] CRAN (R 4.1.2)
914 ##   digest           0.6.30     2022-10-18 [1] CRAN (R 4.1.2)
915 ##   diptest        * 0.76-0     2021-05-04 [1] CRAN (R 4.1.0)
916 ##   distributional   0.3.1      2022-09-02 [1] CRAN (R 4.1.2)
917 ##   dplyr          * 1.0.10     2022-09-01 [1] CRAN (R 4.1.2)
918 ##   DT               0.26       2022-10-19 [1] CRAN (R 4.1.2)
```

```
919  ##   dygraphs          1.1.1.6    2018-07-11 [1] CRAN (R 4.1.0)
920  ##   e1071             1.7-12     2022-10-24 [1] CRAN (R 4.1.2)
921  ##   effectsize        0.8.2      2022-10-31 [1] CRAN (R 4.1.2)
922  ##   ellipse           0.4.3      2022-05-31 [1] CRAN (R 4.1.2)
923  ##   ellipsis          0.3.2      2021-04-29 [1] CRAN (R 4.1.0)
924  ##   emmeans           1.8.2      2022-10-27 [1] CRAN (R 4.1.2)
925  ##   estimability      1.4.1      2022-08-05 [1] CRAN (R 4.1.2)
926  ##   evaluate          0.18       2022-11-07 [1] CRAN (R 4.1.2)
927  ##   extraDistr        1.9.1      2020-09-07 [1] CRAN (R 4.1.0)
928  ##   fansi             1.0.3      2022-03-24 [1] CRAN (R 4.1.2)
929  ##   farver            2.1.1      2022-07-06 [1] CRAN (R 4.1.2)
930  ##   fastmap           1.1.0      2021-01-25 [1] CRAN (R 4.1.0)
931  ##   forcats         * 0.5.2      2022-08-19 [1] CRAN (R 4.1.2)
932  ##   foreach           1.5.2      2022-02-02 [1] CRAN (R 4.1.2)
933  ##   foreign           0.8-83     2022-09-28 [1] CRAN (R 4.1.2)
934  ##   Formula           1.2-4      2020-10-16 [1] CRAN (R 4.1.0)
935  ##   fs                1.5.2      2021-12-08 [1] CRAN (R 4.1.0)
936  ##   furrr             0.3.1      2022-08-15 [1] CRAN (R 4.1.2)
937  ##   future            1.29.0     2022-11-06 [1] CRAN (R 4.1.2)
938  ##   gargle            1.2.1      2022-09-08 [1] CRAN (R 4.1.2)
939  ##   generics          0.1.3      2022-07-05 [1] CRAN (R 4.1.2)
940  ##   gganimate         1.0.8      2022-09-08 [1] CRAN (R 4.1.2)
941  ##   ggdist            3.2.0      2022-07-19 [1] CRAN (R 4.1.2)
942  ##   ggforce           0.4.1      2022-10-04 [1] CRAN (R 4.1.2)
943  ##   ggnewscale        0.4.8      2022-10-06 [1] CRAN (R 4.1.2)
944  ##   ggplot2         * 3.4.0      2022-11-04 [1] CRAN (R 4.1.2)
945  ##   ggpubr            0.5.0      2022-11-16 [1] CRAN (R 4.1.2)
946  ##   ggrepel           0.9.2      2022-11-06 [1] CRAN (R 4.1.2)
947  ##   ggridges          0.5.4      2022-09-26 [1] CRAN (R 4.1.2)
948  ##   ggsignif          0.6.4      2022-10-13 [1] CRAN (R 4.1.2)
```

```
949 ##   ggstance      * 0.3.6      2022-11-16 [1] CRAN (R 4.1.2)
950 ##   gifski          1.6.6-1    2022-04-05 [1] CRAN (R 4.1.2)
951 ##   globals         0.16.2     2022-11-21 [1] CRAN (R 4.1.2)
952 ##   glue            1.6.2      2022-02-24 [1] CRAN (R 4.1.2)
953 ##   googledrive     2.0.0      2021-07-08 [1] CRAN (R 4.1.0)
954 ##   googlesheets4   1.0.1      2022-08-13 [1] CRAN (R 4.1.2)
955 ##   gridExtra       2.3        2017-09-09 [1] CRAN (R 4.1.0)
956 ##   gtable          0.3.1      2022-09-01 [1] CRAN (R 4.1.2)
957 ##   gtools          3.9.4      2022-11-27 [1] CRAN (R 4.1.2)
958 ##   haven           2.5.1      2022-08-22 [1] CRAN (R 4.1.2)
959 ##   HDInterval      0.2.4      2022-11-17 [1] CRAN (R 4.1.2)
960 ##   Hmisc           4.7-2      2022-11-18 [1] CRAN (R 4.1.2)
961 ##   hms             1.1.2      2022-08-19 [1] CRAN (R 4.1.2)
962 ##   htmlTable       2.4.1      2022-07-07 [1] CRAN (R 4.1.2)
963 ##   htmltools       0.5.3      2022-07-18 [1] CRAN (R 4.1.2)
964 ##   htmlwidgets     1.5.4      2021-09-08 [1] CRAN (R 4.1.0)
965 ##   httpuv          1.6.6      2022-09-08 [1] CRAN (R 4.1.2)
966 ##   httr            1.4.4      2022-08-17 [1] CRAN (R 4.1.2)
967 ##   igraph          1.3.5      2022-09-22 [1] CRAN (R 4.1.2)
968 ##   inline          0.3.19     2021-05-31 [1] CRAN (R 4.1.2)
969 ##   insight         0.18.8     2022-11-24 [1] CRAN (R 4.1.2)
970 ##   interp          1.1-3      2022-07-13 [1] CRAN (R 4.1.2)
971 ##   iterators       1.0.14     2022-02-05 [1] CRAN (R 4.1.2)
972 ##   jpeg            0.1-9      2021-07-24 [1] CRAN (R 4.1.0)
973 ##   jsonlite        1.8.3      2022-10-21 [1] CRAN (R 4.1.2)
974 ##   kableExtra    * 1.3.4      2021-02-20 [1] CRAN (R 4.1.2)
975 ##   KernSmooth      2.23-20    2021-05-03 [1] CRAN (R 4.1.3)
976 ##   knitr           1.41       2022-11-18 [1] CRAN (R 4.1.2)
977 ##   labeling        0.4.2      2020-10-20 [1] CRAN (R 4.1.0)
978 ##   LaplacesDemon   16.1.6     2021-07-09 [1] CRAN (R 4.1.0)
```

```
979   ##   later             1.3.0      2021-08-18 [1] CRAN (R 4.1.0)
980   ##   latexdiffr      * 0.1.0      2021-05-03 [1] CRAN (R 4.1.0)
981   ##   lattice           0.20-45    2021-09-22 [1] CRAN (R 4.1.3)
982   ##   latticeExtra      0.6-30     2022-07-04 [1] CRAN (R 4.1.2)
983   ##   lazyeval          0.2.2      2019-03-15 [1] CRAN (R 4.1.0)
984   ##   lifecycle         1.0.3      2022-10-07 [1] CRAN (R 4.1.2)
985   ##   linguisticsdown * 1.2.0      2019-03-01 [1] CRAN (R 4.1.0)
986   ##   listenv           0.8.0      2019-12-05 [1] CRAN (R 4.1.0)
987   ##   lme4            * 1.1-31     2022-11-01 [1] CRAN (R 4.1.2)
988   ##   lmerTest          3.1-3      2020-10-23 [1] CRAN (R 4.1.0)
989   ##   loo               2.5.1      2022-03-24 [1] CRAN (R 4.1.2)
990   ##   lpSolve           5.6.17     2022-10-10 [1] CRAN (R 4.1.2)
991   ##   lubridate         1.9.0      2022-11-06 [1] CRAN (R 4.1.2)
992   ##   magick          * 2.7.3      2021-08-18 [1] CRAN (R 4.1.0)
993   ##   magrittr        * 2.0.3      2022-03-30 [1] CRAN (R 4.1.2)
994   ##   markdown          1.4        2022-11-16 [1] CRAN (R 4.1.2)
995   ##   MASS              7.3-58.1   2022-08-03 [1] CRAN (R 4.1.2)
996   ##   Matrix          * 1.5-1      2022-09-13 [1] CRAN (R 4.1.2)
997   ##   matrixStats       0.63.0     2022-11-18 [1] CRAN (R 4.1.2)
998   ##   memoise           2.0.1      2021-11-26 [1] CRAN (R 4.1.0)
999   ##   mime              0.12       2021-09-28 [1] CRAN (R 4.1.0)
1000  ##   miniUI            0.1.1.1    2018-05-18 [1] CRAN (R 4.1.0)
1001  ##   minqa             1.2.5      2022-10-19 [1] CRAN (R 4.1.2)
1002  ##   modelr            0.1.10     2022-11-11 [1] CRAN (R 4.1.2)
1003  ##   multcomp          1.4-20     2022-08-07 [1] CRAN (R 4.1.2)
1004  ##   munsell           0.5.0      2018-06-12 [1] CRAN (R 4.1.0)
1005  ##   MVBeliefUpdatr  * 0.0.1.0002 2022-11-30 [1] Github (hlplab/MVBeliefUpdatr@5972af5)
1006  ##   mvtnorm           1.1-3      2021-10-08 [1] CRAN (R 4.1.0)
1007  ##   nlme              3.1-160    2022-10-10 [1] CRAN (R 4.1.2)
1008  ##   nloptr            2.0.3      2022-05-26 [1] CRAN (R 4.1.2)
```

```
1009 ##   nnet          7.3-18      2022-09-28 [1] CRAN (R 4.1.2)
1010 ##   numDeriv      2016.8-1.1  2019-06-06 [1] CRAN (R 4.1.0)
1011 ##   pander        0.6.5       2022-03-18 [1] CRAN (R 4.1.2)
1012 ##   papaja      * 0.1.1.9001  2022-11-30 [1] Github (crsh/papaja@3b1face)
1013 ##   parallelly    1.32.1      2022-07-21 [1] CRAN (R 4.1.2)
1014 ##   parameters    0.20.0      2022-11-21 [1] CRAN (R 4.1.2)
1015 ##   patchwork   * 1.1.2       2022-08-19 [1] CRAN (R 4.1.2)
1016 ##   phonR       * 1.0-7       2016-08-25 [1] CRAN (R 4.1.0)
1017 ##   pillar        1.8.1       2022-08-19 [1] CRAN (R 4.1.2)
1018 ##   pkgbuild      1.4.0       2022-11-27 [1] CRAN (R 4.1.2)
1019 ##   pkgconfig     2.0.3       2019-09-22 [1] CRAN (R 4.1.0)
1020 ##   pkgload       1.3.2       2022-11-16 [1] CRAN (R 4.1.2)
1021 ##   plotly        4.10.1      2022-11-07 [1] CRAN (R 4.1.2)
1022 ##   plyr          1.8.8       2022-11-11 [1] CRAN (R 4.1.2)
1023 ##   png           0.1-8       2022-11-29 [1] CRAN (R 4.1.3)
1024 ##   polyclip      1.10-4      2022-10-20 [1] CRAN (R 4.1.2)
1025 ##   posterior   * 1.3.1       2022-09-06 [1] CRAN (R 4.1.2)
1026 ##   prettyunits   1.1.1       2020-01-24 [1] CRAN (R 4.1.0)
1027 ##   processx      3.8.0       2022-10-26 [1] CRAN (R 4.1.2)
1028 ##   profvis       0.3.7       2020-11-02 [1] CRAN (R 4.1.0)
1029 ##   progress      1.2.2       2019-05-16 [1] CRAN (R 4.1.0)
1030 ##   promises      1.2.0.1     2021-02-11 [1] CRAN (R 4.1.0)
1031 ##   proxy         0.4-27      2022-06-09 [1] CRAN (R 4.1.2)
1032 ##   ps            1.7.2       2022-10-26 [1] CRAN (R 4.1.2)
1033 ##   purrr       * 0.3.5       2022-10-06 [1] CRAN (R 4.1.2)
1034 ##   R6            2.5.1       2021-08-19 [1] CRAN (R 4.1.0)
1035 ##   rbibutils     2.2.10      2022-11-15 [1] CRAN (R 4.1.2)
1036 ##   RColorBrewer  1.1-3       2022-04-03 [1] CRAN (R 4.1.2)
1037 ##   Rcpp        * 1.0.9       2022-07-08 [1] CRAN (R 4.1.2)
1038 ##   RcppParallel  5.1.5       2022-01-05 [1] CRAN (R 4.1.2)
```

```
1039  ##  Rdpack          2.4        2022-07-20 [1] CRAN (R 4.1.2)
1040  ##  readr         * 2.1.3      2022-10-01 [1] CRAN (R 4.1.2)
1041  ##  readxl          1.4.1      2022-08-17 [1] CRAN (R 4.1.2)
1042  ##  remotes         2.4.2      2021-11-30 [1] CRAN (R 4.1.0)
1043  ##  reprex          2.0.2      2022-08-17 [1] CRAN (R 4.1.2)
1044  ##  reshape2        1.4.4      2020-04-09 [1] CRAN (R 4.1.0)
1045  ##  rlang         * 1.0.6      2022-09-24 [1] CRAN (R 4.1.2)
1046  ##  rmarkdown       2.18       2022-11-09 [1] CRAN (R 4.1.2)
1047  ##  rpart           4.1.19     2022-10-21 [1] CRAN (R 4.1.2)
1048  ##  rsample       * 1.1.1      2022-12-07 [1] CRAN (R 4.1.2)
1049  ##  rstan           2.21.7     2022-09-08 [1] CRAN (R 4.1.2)
1050  ##  rstantools      2.2.0      2022-04-08 [1] CRAN (R 4.1.2)
1051  ##  rstatix         0.7.1      2022-11-09 [1] CRAN (R 4.1.2)
1052  ##  rstudioapi      0.14       2022-08-22 [1] CRAN (R 4.1.2)
1053  ##  rvest           1.0.3      2022-08-19 [1] CRAN (R 4.1.2)
1054  ##  sandwich        3.0-2      2022-06-15 [1] CRAN (R 4.1.2)
1055  ##  scales          1.2.1      2022-08-20 [1] CRAN (R 4.1.2)
1056  ##  sessioninfo     1.2.2      2021-12-06 [1] CRAN (R 4.1.0)
1057  ##  sf              1.0-9      2022-11-08 [1] CRAN (R 4.1.2)
1058  ##  shiny           1.7.3      2022-10-25 [1] CRAN (R 4.1.2)
1059  ##  shinyjs         2.1.0      2021-12-23 [1] CRAN (R 4.1.0)
1060  ##  shinystan       2.6.0      2022-03-03 [1] CRAN (R 4.1.2)
1061  ##  shinythemes     1.2.0      2021-01-25 [1] CRAN (R 4.1.0)
1062  ##  StanHeaders     2.21.0-7   2020-12-17 [1] CRAN (R 4.1.0)
1063  ##  stringi         1.7.8      2022-07-11 [1] CRAN (R 4.1.2)
1064  ##  stringr       * 1.4.1      2022-08-20 [1] CRAN (R 4.1.2)
1065  ##  survival        3.4-0      2022-08-09 [1] CRAN (R 4.1.2)
1066  ##  svglite         2.1.0      2022-02-03 [1] CRAN (R 4.1.2)
1067  ##  svUnit          1.0.6      2021-04-19 [1] CRAN (R 4.1.0)
1068  ##  systemfonts     1.0.4      2022-02-11 [1] CRAN (R 4.1.2)
```

```
1069 ##   tensorA        0.36.2   2020-11-19 [1] CRAN (R 4.1.0)
1070 ##   TH.data        1.1-1    2022-04-26 [1] CRAN (R 4.1.2)
1071 ##   threejs        0.3.3    2020-01-21 [1] CRAN (R 4.1.0)
1072 ##   tibble       * 3.1.8    2022-07-22 [1] CRAN (R 4.1.2)
1073 ##   tidybayes    * 3.0.2    2022-01-05 [1] CRAN (R 4.1.2)
1074 ##   tidyr        * 1.2.1    2022-09-08 [1] CRAN (R 4.1.2)
1075 ##   tidyselect     1.2.0    2022-10-10 [1] CRAN (R 4.1.2)
1076 ##   tidyverse    * 1.3.2    2022-07-18 [1] CRAN (R 4.1.2)
1077 ##   timechange     0.1.1    2022-11-04 [1] CRAN (R 4.1.2)
1078 ##   tinylabels   * 0.2.3    2022-02-06 [1] CRAN (R 4.1.2)
1079 ##   transformr     0.1.4    2022-08-18 [1] CRAN (R 4.1.2)
1080 ##   tufte          0.12     2022-01-27 [1] CRAN (R 4.1.2)
1081 ##   tweenr         2.0.2    2022-09-06 [1] CRAN (R 4.1.2)
1082 ##   tzdb           0.3.0    2022-03-28 [1] CRAN (R 4.1.2)
1083 ##   units          0.8-0    2022-02-05 [1] CRAN (R 4.1.2)
1084 ##   urlchecker     1.0.1    2021-11-30 [1] CRAN (R 4.1.0)
1085 ##   usethis        2.1.6    2022-05-25 [1] CRAN (R 4.1.2)
1086 ##   utf8           1.2.2    2021-07-24 [1] CRAN (R 4.1.0)
1087 ##   vctrs          0.5.1    2022-11-16 [1] CRAN (R 4.1.2)
1088 ##   viridis        0.6.2    2021-10-13 [1] CRAN (R 4.1.0)
1089 ##   viridisLite    0.4.1    2022-08-22 [1] CRAN (R 4.1.2)
1090 ##   vroom          1.6.0    2022-09-30 [1] CRAN (R 4.1.2)
1091 ##   webshot        0.5.4    2022-09-26 [1] CRAN (R 4.1.2)
1092 ##   withr          2.5.0    2022-03-03 [1] CRAN (R 4.1.2)
1093 ##   xfun           0.35     2022-11-16 [1] CRAN (R 4.1.2)
1094 ##   xml2           1.3.3    2021-11-30 [1] CRAN (R 4.1.0)
1095 ##   xtable         1.8-4    2019-04-21 [1] CRAN (R 4.1.0)
1096 ##   xts            0.12.2   2022-10-16 [1] CRAN (R 4.1.2)
1097 ##   yaml           2.3.6    2022-10-18 [1] CRAN (R 4.1.2)
1098 ##   zoo            1.8-11   2022-09-17 [1] CRAN (R 4.1.2)
```

```
1099  ##
1100  ##  [1] /Library/Frameworks/R.framework/Versions/4.1/Resources/library
1101  ##
1102  ## ----------------------------------------------------------------------
```