THE EFFECTS OF ATTENTION AND PRIOR KNOWLEDGE ON PERCEPTION AND MISPERCEPTION OF SPEECH

ALINA KUIMOVA

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN LINGUISTICS

YORK UNIVERSITY

TORONTO, ONTARIO

August 2022

Abstract

Misperceptions are common in everyday conversation. Previous work shows that misperception derives from a weak neural representation of sounds that deviate from prior expectations (prediction error). Attention enhances the encoding of prediction error and supports speech perception in challenging listening conditions, suggesting that increased attentional engagement might reduce the rate of misperceptions driven by plausible yet misinformative expectations. We induced frequent misperception in a word discrimination task with degraded spoken words preceded by matching, mismatching, and partially mismatching written text, using monetary incentives to manipulate listeners' attention. Contrary to our predictions, incentives increased misperception on partial mismatch trials but improved perceptual accuracy on match trials. Pupillometry showed that incentives loaded both proactive and reactive control, suggesting increased involvement of top-down predictive processes. We conclude that higher attentional engagement increases reliance on prior knowledge when sensory detail is insufficient, which only exacerbates prediction-induced mishearing—at least in word discrimination tasks.

Keywords: pupillometry, misperception, prior expectations, attention, motivation, reward, incentives, speech perception

Acknowledgments

First and foremost, I would like to express the deepest of gratitude to my supervisor, Chandan Narayan. I couldn’t have asked for better support with every aspect of this project. Thank you for making this work possible, bearing with my often-unyielding stubbornness and always being there for me during this challenging year.

Matt Davis, you have my most profound appreciation and utmost respect. I can’t thank you enough for all the feedback and advice that you gave me throughout the year—for I needed it more than anything, more than I even realized.

To my thesis committee, Scott Adler and Liz Smeets, and other scientists who have contributed to this project—thank you for challenging me and finding ways to make this work better.

I would also like to thank my department, and particularly Geoff Lawrence and Rose Frezza-Edgecombe, for letting me challenge myself and keeping my spirits high when the initial confidence wore off. A special thank you goes to Gabriela Alboiu, Philippe Bourdin, and Chris Green for (un)knowingly helping me go through this uneasy transition and for being the wonderful human beings that you are.

Last but not least, I want to thank my friends and family—I wouldn’t have completed this thesis without you always having my back. Thank you to my stoic cat Marcus, whose unrelenting meows reminded me of life behind my screen.

Finally, Dhyan, I would have never got where I am now without you. Thank you for crossing my path and changing my life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Perception and misperception of speech

1.1.1 The effects of prior knowledge in speech perception

Perception is an active process. Predictive coding, a major theory of brain function (Friston, 2005; Rao & Ballard, 1999), posits that our perceptual experiences are the result of Bayesian inference—our brain's attempt to predict what happens next given what we already know. As we go about life being exposed to a myriad of sounds, our brain constructs a predictive model of expected sensory inputs based on the knowledge of past events (Clark, 2013; Friston, 2005). As these sensory signals unfold, model predictions—"virtual" inputs generated by the expected causes—are compared to the sensory evidence. The difference between predictions and observations—the prediction error—is calculated at each processing stage and propagated to the higher levels of the cortical hierarchy. These prediction errors are then used to update perceptual hypotheses. In this way, the predictive model adapts to the incoming sensory inputs, constantly fine-tuning itself so that it can make better predictions in the future. Perception is thus effectively achieved by minimizing the error between the predicted and the observed, the known and the seen, the expected and the heard (Feldman & Friston, 2010; Friston, 2010; Hohwy, 2012). Crucially, prior knowledge and expectations are pivotal for perceptual inference (de Lange et al., 2018; den Ouden et al., 2012).

Speech perception is no different: listeners routinely make use of prior expectations to help constrain perceptual interpretation (Davis & Johnsrude, 2007; Kuperberg & Jaeger, 2016). Preceding sentence context is used to predict upcoming words: we get surprised if an utterance like "the day was breezy so the boy went outside to fly a…" suddenly ends with a word "plane", expecting to hear something more like "kite". As in more general cases of perception, violation of such contextual expectations evokes a strong neural response to mismatch (DeLong et al., 2005). The degree of neural surprisal depends on the certainty of current predictions: violation of highly constraining contexts (e.g., “to fly a... plane”) triggers additional neural processing, above and beyond the normal response to surprise evoked by less constraining contexts (“the boy went outside and *saw* a… plane”) (DeLong et al., 2005; Kutas & Hillyard, 1984). On the other hand, expected, semantically-predictable words, such as “kite” in the sentence above, tend to be processed faster and elicit lower neural activity (DeLong et al., 2005; Kutas & Hillyard, 1980, 1984), consistent with the “silencing” effect of prediction (Feldman & Friston, 2010; Friston, 2005).

While most of the time we are barely conscious of making perceptual inferences during listening, facilitatory effects of prediction become transparent when listening conditions degrade. When the sensory inputs are ambiguous or unreliable—due to acoustic distortion or hearing impairment,—informative prior knowledge enhances speech clarity (Signoret et al., 2018; Sohoglu et al., 2012), improves speech comprehension (Corps & Rabagliati, 2020; Davis et al., 2005; Obleser et al., 2007; Remez et al., 1981), reduces listening effort (Winn, 2016), and facilitates perceptual learning of barely intelligible speech (Davis et al., 2005). One way to induce strong prior expectations with degraded speech is identity priming—revealing the identity of the degraded sentence by presenting listeners with its written or clearly-spoken version just before its auditory replay. In cases of severely degraded speech, knowing the content of the distorted utterance prior to hearing it completely transforms the perceptual experience, leading to a perceptual “pop-out” in which a previously incomprehensible speech becomes perfectly intelligible (Davis et al., 2005; Hervais-Adelman et al., 2008). The perceptual effect of informative prior knowledge is akin to hearing an acoustically clearer speech, yet their neural signatures could not be more different. Clearer speech increases neural activity in peri-auditory regions—matching written text, on the contrary, decreases it (Blank & Davis, 2016; Sohoglu et al., 2012; Sohoglu & Davis, 2016).

1.1.2 Misperception as overreliance on prior knowledge

Informative prior knowledge improves speech perception both in the short term (perceptual inference) and in the long run (perceptual learning) (Sohoglu & Davis, 2016). *Mis*informative prior knowledge (such as mismatching written primes), on the other hand, can further reduce the subjective intelligibility of an already degraded speech (Signoret et al., 2018; Sohoglu et al., 2012, 2014; Wild, Davis, et al., 2012). Unsurprisingly, when prior expectations conflict with the sensory evidence, listeners' word recognition performance tends to suffer (Sohoglu et al., 2012, 2014). Oftentimes, however, our prior expectations neither fully match nor fully mismatch reality. In ideal listening conditions, a clear, high-precision sensory signal easily overrides a misinformative prior and ensures veridical perception. When acoustics is poor, though—be it due to hearing deficits, noisy background, or mechanical distortion—the precision of the sensory signal decreases due to masking or the loss of spectro-temporal detail. Given a low sensory precision, any surprise about the mismatch between prior expectations and the sensory evidence is often correspondingly low, which may lead to an erroneous confirmation of the prior. In its turn, a failure to adjust or reject strong yet erroneous predictions ensues auditory illusions (Blank et al., 2018).

One classic and ubiquitous example is misheard lyrics (Bond, 2005). In most cases, this type of misperception is the outcome of relatively low-precision acoustics: together, background music and atypical pronunciation often create a strong basis for ambiguity and perceptual confusion. This explains why the chorus of R.E.M.'s *The Sidewinder Sleeps Tonite*—"Call me when you try to wake her up"—is easier perceived as "calling Jamaica" rather than the original refrain. However, misperceptions can be just as easily triggered by contextual cues. Presenting erroneous yet plausible captions—or other types of visual context—along with almost any song is a reliable way to induce a strong perceptual bias that can lead to misperception of a target passage (Beck Lidén et al., 2016). Eric Carmen's *All by myself* accompanied by a playful animation of a gnome sitting on the shoulder of an American ex-president stubbornly triggers "Obama's elf, don't wanna be… Obama's elf, anymore"[1]. Interestingly, such perceptual illusions may persist even when the original lyrics are known. Perceptual ambiguity simply leaves both interpretations of the signal plausible (Beck Lidén et al., 2016).

While misperceptions are more common when speech clarity is reduced in one way or another, the famous McGurk illusion demonstrates that prior expectations can affect the perception of even clearly spoken stimuli. Hearing a person say [ba] while seeing his mouth pronounce [ga] leads one to believe in hearing [da] (McGurk & MacDonald, 1976). A partial phonetic overlap between expected and sensory inputs further increases the likelihood of misperception and the strength of the auditory illusion (Blank et al., 2018; Sohoglu et al., 2014). Naturally-occurring "slips of the ear" also tend to be phonetically close to the target: the mean edit distance between the "slip" and the target is three phonemes (Felty et al., 2013). Misperceptions of isolated words most often involve mishearing and substitution of syllables, or deletions and additions of individual segments (Felty et al., 2013). In the context of sentences, slips of the ear tend to be a combination of misheard individual words and a subsequent "rationalization" of these initial perceptual errors (Winn & Teece, 2021). Indeed, as the speech

[1] A reader is invited to personally attest these effects: https://www.youtube.com/shorts/6mfFHWIYPKM

unfolds, a single misperceived item may cause a strong semantic incoherence with upcoming words. An attempt to resolve this ambiguity may trigger a re-analysis of other parts of the sentence, transforming the original utterance—e.g., “she made the bed with clean sheets”—into something entirely different, such as “she made the bagel with cream cheese” (Winn & Teece, 2021).

All these factors—poor acoustics, auditory deficits, and erroneous yet plausible prior expectations—make speech misperceptions more common. Frequent misperceptions may result in a failure of speech comprehension, breakdowns in communication, and, in extreme cases, even social withdrawal—with severe consequences for cognition, well-being, and quality of life (Pichora-Fuller et al., 2015). And while sentence context can often disambiguate misheard words, acoustic distortion and hearing impairment slow down the processing of contextual cues, particularly in natural speech, when the misperceived sentence is immediately followed by another utterance (Winn, 2016; Winn & Moore, 2018). The speed of speech processing tends to be particularly disrupted in individuals with sensorineural hearing loss and cochlear implant users (Winn, 2016)—the exact category of listeners who heavily rely on contextual information to offset reduced audibility (Dingemanse & Goedegebure, 2019; Signoret & Rudner, 2019). These listeners may “repair” misperceptions at the end, but their inefficient use of contextual cues coupled with the already compromised auditory encoding makes speech comprehension effortful and leads to listening fatigue—a common complaint of people with hearing aids (Winn & Moore, 2018). Chronically elevated listening effort has severe health consequences in itself: it has been associated with cortical thinning and gray matter loss in the prefrontal brain regions (Rosemann & Thiel, 2020). These two, in turn, have been linked to the development and progression of dementia (Bakkour et al., 2009; Dickerson et al., 2009; Zarei et al., 2013), a neurodegenerative disease that tends to be predominant among hearing-compromised individuals (Gurgel et al., 2014; Lin et al., 2013). It is therefore important to study factors that affect the frequency of perceptual confusion in speech in order to guide hearing rehabilitation approaches to facilitating communication repairs and reducing listening effort for people with hearing loss.

1.1.3 Prediction error as a neural index of (mis)perception

The result of poor acoustics and misinformative prior expectations, misperceptions can be easily induced in the lab (Beck Lidén et al., 2016; Blank et al., 2018; Sohoglu et al., 2014). Blank and colleagues (2018) presented listeners with spoken noise-vocoded words preceded by written words that either matched, partially mismatched, or completely mismatched the acoustics. Partial phonetic overlap between a prior (induced by written text) and a low-fidelity sensory signal resulted in a high percentage of misperceptions. On about 40% of partial mismatch trials, listeners mistakenly reported that the distorted word they heard (e.g., /pIt/ 'pit') was identical to one they saw (e.g., PICK). An fMRI analysis revealed that the overall magnitude of the BOLD signal in the bilateral superior temporal sulcus (STS) increased when listeners correctly perceived the mismatch between written and spoken words. Total mismatch pairs and detected partial mismatch pairs evoked elevated neural responses of a similar magnitude. Multivoxel pattern analysis decoded this increased activity as reflecting enhanced neural representations of mismatching (i.e., -/k/, +/t/)—rather than common (/pI/)—sounds. This pattern of activity suggests that a stronger neural representation of mismatch (i.e., the prediction error) causes larger updates to the sensory prediction, leading to a timely rejection of the prior and ensuing veridical perception. This interpretation is supported by the fact that word pairs perceived as "same", both correctly (matching pairs) and incorrectly (partial mismatch pairs), elicited equally suppressed neural activity, reflecting the “silencing” effect of confirmed expectations (Feldman & Friston, 2010; Friston, 2005).

Thus, the strength of prediction error distinguishes veridical and erroneous perception of speech: weak neural representation of mismatching sounds results in misperception, while a stronger representation of mismatch leads to more accurate perceptual outcomes. However, the strength of prediction error might depend on multiple factors. One possibility is that a stronger prediction error signal was due to purely physical traits of the stimuli, such as higher acoustic dissimilarity between written and spoken words. That is, some written/spoken pairs were correctly perceived as “different” more often because they were more acoustically dissimilar than other pairs. Blank and colleagues (2018) indeed report that acoustic dissimilarity strongly correlated with the rate of accurate perception on partial mismatch trials. The nature of sounds that distinguish written and spoken words (e.g., -k/+p for the *kit-pit* pair), in particular, was predictive of the likelihood of misperception—more so than the nature of common sounds (e.g., /_it/ for the *kit-pit* pair). There is, however, another possibility: perceptual outcomes could be determined by dynamic shifts in attentional engagement during listening, including momentary lapses of attention.

Under the predictive coding framework, prediction errors reflect not only the content of sensory inputs but also the level of uncertainty about predictions, or “inverse precision” (Feldman & Friston, 2010; Friston, 2010; Hohwy, 2012). Sensory evidence is weighed according to the precision of prediction error: a stronger mismatch signal reflects greater reliability of sensory inputs—a weaker one signals uncertainty. A recent version of the predictive coding theory assumes that this function of “scaling” prediction error is taken on by attention (Feldman & Friston, 2010; Friston, 2010). Attention optimizes the expected precision of sensory predictions by increasing the synaptic gain of prediction error units (Feldman & Friston, 2010). This attentional scaling, in turn, leads to a heightened selectivity for the attended information and, by consequence, a stronger neural response to mismatch (Auksztulewicz & Friston, 2015). Since stronger prediction errors were linked to veridical perception, this model predicts that increased attentional engagement could strengthen the neural representation of mismatch and thus decrease the rate of expectation-induced misperceptions. Momentary lapses of attention or even longer periods of mind wandering, on the other hand, are likely to induce more frequent misperceptions. Perceptual outcomes could thus be determined by the level of attentional engagement during listening rather than solely by the acoustic (dis)similarity of predictions and sensory inputs. The present thesis sets out to investigate this possibility.

1.2. Attention and cognitive control in speech perception

1.2.1 Attention and its effects on speech processing

Speech perception critically depends on attention. Yet, our attention to the external environment tends to wax and wane over time. Not only is mind wandering recognized as the brain's default mode of operation (Mason et al., 2007), but people may pay no attention to what they are doing about 50% of the time—even when purportedly "on task" (Smallwood et al., 2008). This general pattern endures during listening and other language-related tasks. For instance, Boudewyn & Carter (2018), who combined the probe-caught method—a common index of mind wandering—with an ecologically valid listening task, report that people find themselves "zoned out" about 30% of the time while listening to a story. Another listening study investigated attentional dynamics in a multispeaker context using eye tracking. It found that participants shift their gaze—together with their locus of attention—from the target speaker to other places in the environment, including an interfering talker, for over 10% of the time (Shavit-Cohen & Zion Golumbic, 2019). Predictably, withdrawing one's attention, even momentarily, has negative consequences for speech recognition and language comprehension performance. Both studies report that brief lapses of attention negatively affect speech comprehension since listeners are more likely to miss key information in the target speech stream while attending elsewhere (Boudewyn & Carter, 2018; Shavit-Cohen & Zion Golumbic, 2019).

Yet another, perhaps more extreme, piece of evidence comes from a pharmacological neuroimaging study by Davis and colleagues (2007) who empirically demonstrated that language processing in the brain is impaired at reduced levels of awareness. Instead of observing the consequences of the naturally occurring "ebb and flow" of attention, this study investigated what happens when listeners are *physically unable* to attend to speech due to sedation. They found that neural responses in the bilateral temporal lobe remained robust regardless of the level of sedation, suggesting that lower-level, perceptual processes stay relatively intact even at the minimal level of conscious awareness. Reduced neural activity in the inferior frontal and posterior temporal areas, however, indicated that this is not the case for higher-order language processes, such as comprehension and memory for speech, which start to suffer at the lightest level of sedation and come to a halt at deep sedation. This fMRI evidence was corroborated by behavioural report, as listeners were unable to recall sentences they heard while deeply sedated. Thus, when people fail to sustain their attention during listening—be it because of mind wandering or extreme drowsiness—their speech processing capacity is reduced.

1.2.2 Attention and effortful listening

Attention becomes even more important as listening conditions degrade. Wild and colleagues (2012) directly compared the perception of clear and distorted speech under distraction. In this study, memory for clearly spoken sentences remained sharp even when listeners were distracted by a parallel task, while recognition and recall of degraded speech strongly depended on attention. Distorted speech was highly intelligible when attended, but processing and subsequent recognition of distorted sentences decreased as a function of intelligibility when listeners' attention was directed elsewhere. This attentional modulation of perception was reflected in neural activity. Brain responses to degraded speech along the STS were enhanced under direct attention, irrespective of intelligibility. When listeners were distracted, however, STS activity correlates with (self-reported) intelligibility, demonstrating that speech-selective temporal regions lose speech sensitivity when the auditory input is neglected or ignored (see also Sabri et al., 2008; Ritz et al., 2021). Frontal regions, including the cingulo-opercular network and left inferior frontal gyrus (LIFG), showed an elevated response to degraded compared to clear speech, but engaged only when this speech was attended (Wild, Yusuf, et al., 2012). This noise-elevated response in putative attentional networks suggests that attentional control systems engage to support speech perception in a top-down fashion when listening conditions degrade.

Cingulo-opercular network (CON), in particular, comprised of anterior cingulate cortex and anterior insula, frequently exhibits increased neural response when speech comprehension becomes effortful (Adank et al., 2012; Alain et al., 2018; Erb & Obleser, 2013; Hervais-Adelman et al., 2012; Ritz et al., 2021; Vaden et al., 2013). CON is counted as one of the task-positive attentional networks, responsible for attentional monitoring, tonic alertness (i.e., mentally effortful, endogenously driven vigilance), and optimization of performance on challenging tasks (Dosenbach et al., 2008; Kerns et al., 2004; Sadaghiani & D'Esposito, 2015; Weissman et al., 2005). Cingulo-opercular activity and connectivity increase in response to growing task demands—often upon error detection—when it is clear that optimal performance calls for sustained cognitive control. CO network is also thought to be responsible for redirecting attention back to the task after periods of brief disengagement (Eichele et al., 2008). During listening, elevated cingulo-opercular activity was attested both in response to a higher cognitive load (due to a parallel task) and in response to an increased listening effort (e.g., due to degraded speech with lower acoustic detail), indicating a domain-general rather than a language-specific function (Ritz et al., 2021).

In speech perception tasks, elevated cingulo-opercular activity has been associated with better recognition of degraded speech in younger and older adults, both with and without hearing loss (Erb & Obleser, 2013; Vaden et al., 2013, 2015, 2016). The magnitude of CO activity can predict word recognition performance on subsequent trials: elevated CO activity increases the likelihood of accurate perception—low activity predicts an impeding perceptual difficulty (Vaden et al., 2013, 2015, 2016). Interestingly, the degree to which CO engagement predicts speech recognition performance in older listeners is determined not by the severity of hearing loss—which directly affects the quality of the incoming sensory input,—but rather by the age-related declines in cognitive function, including attention (Vaden et al., 2015). Relatedly, elevated cingulo-opercular activity anticipates the subsequent engagement of the frontoparietal network—a functional system responsible for phasic alertness, that adaptively adjusts attentional control on a case-by-case basis (Dosenbach et al., 2008; Kerns et al., 2004). In challenging listening conditions, the frontoparietal regions are thought to support speech comprehension through the top-down use of higher-order linguistic information, such as semantic and contextual cues (Peelle, 2018; Smirnov et al., 2014). The extent to which these domain-general intentional and attentional networks engage during effortful language processing often determines perceptual outcomes of listening (Obleser et al., 2007; Ritz et al., 2021; Rysop et al., 2021; Vaden et al., 2013).

Attention proved critical for other top-down processes, such as perceptual learning for speech. Perception of degraded, yet still intelligible, speech rapidly improves with exposure, even if the task involves no more than passive listening (Davis et al., 2005; Hervais-Adelman et al., 2011). Yet, no perceptual learning occurs when listeners actively attend elsewhere, e.g., to a competing auditory or visual stream (Huyck & Johnsrude, 2012). Interestingly, these more robust, longer-lasting changes in performance associated with perceptual learning appear to be driven by the same neural mechanism—the minimization of prediction error—as the more immediate effects of prior knowledge, i.e., the aforementioned perceptual "pop-out" (Sohoglu & Davis, 2016). Both prior knowledge and perceptual learning lead to reduced neural responses in a region of the superior temporal gyrus (STG). The magnitude of this reduction for prior knowledge effects predicts the degree to which STG activity drops once perceptual training is complete and correlates with achieved behavioural improvements. Crucially, if perceptual learning, being the end result of prediction error minimization, critically depends on attention, then attention must play a key role in the computation of prediction error for the accurate perceptual inference "online".

1.2.3 Neuroeconomics of cognitive control

As the sensory precision of speech declines with distortion, listeners experience an increased need to suppress background noise, selectively focus attention on relevant acoustic cues, and inhibit competition from phonetically-similar lexical alternatives. However, the escalation of listening demand alone does not warrant an automatic (and prolonged) engagement of domain-general attentional control systems. To benefit from upregulated cognitive control, listeners must actively attend to the speech stream in question (Ritz et al., 2021; Wild, Yusuf, et al., 2012). They must also command sufficient cognitive resources to support a high listening load: if attention is divided between competing tasks, for example, the processing of degraded speech will be blocked by parallel task demands (Ritz et al., 2021). The intelligibility of speech also matters. Cognitive control tends to be at its highest at the moderate levels of speech clarity: if speech is too easy or too hard to comprehend, control is withdrawn, and activity in frontal and CO networks dwindles (Obleser et al., 2007; Poldrack et al., 2001; Rysop et al., 2021; Zekveld et al., 2006). This inverted U-shaped response reflects adaptive control. This mechanism, much like a thermostat, activates in response to decreasing task performance and allocates additional cognitive resources to bring it back to acceptable levels—as long as the task utility is sufficiently high.

Indeed, cognitive control is effortful, and the benefit of engaging additional resources must outweigh the costs (Botvinick & Braver, 2015; Shenhav et al., 2017; Westbrook & Braver, 2015). In the case of listening-in-noise, these costs manifest as a feeling of increasing effort and fatigue—a frequent complaint of listeners impacted by hearing loss (Eckert et al., 2016; McGarrigle et al., 2017; Pichora-Fuller et al., 2016). The cingulo-opercular network, implicated in adaptive control, is thought to perform such cost-benefit analysis (Aston-Jones & Cohen, 2005; Peelle, 2018; Shenhav et al., 2013). Cingulate neurons weigh the benefit arising from the improved task performance against its expected effort-related costs (Holroyd & McClure, 2015). They then signal the frontoparietal network to implement control on high-utility tasks (Eckert et al., 2016). In listening tasks, utility is determined by the expected benefits of successful comprehension relative to the effort required to achieve this level of performance. An illustrative example comes from Eckert and colleagues (2016) who modeled how the relative value of speech recognition would change depending on speech content: listening to one’s grandchildren is more rewarding an activity than listening to a documentary about lint. Despite equal speech intelligibility, participants in a more valuable listening condition are expected to sustain increased levels of cingulo-opercular activity for a longer period of time before experiencing listening fatigue.

The informational or emotional value of speech determines both the degree of adaptive gain—i.e., the flexible adjustment of control for the purpose of improving recognition performance—and listeners’ ability to maintain higher attentional engagement (Eckert et al., 2016; Herrmann & Johnsrude, 2020; Peelle, 2018; Pichora-Fuller et al., 2016). External sources of motivation, such as a monetary reward, can also modulate the perceived utility of listening tasks and thereby affect cognitive control (Carolan et al., 2021; Koelewijn et al., 2018; Plain et al., 2021; Richter, 2016). Performance-contingent rewards increase arousal, enhance the level of alertness and decrease the likelihood of attentional lapses (Esterman et al., 2014). Reward has clear effects on listening engagement, particularly in challenging tasks: participants were shown to exert the highest levels of effort and adaptive control when completing a high-difficulty/high-reward listening task compared to high-difficulty/low-reward or low-difficulty/high-reward tasks (Koelewijn et al., 2021; Richter, 2016).

It is worth noting, however, that the effects of reward on perceptual outcomes of listening are less clear. In some studies, when a reward is at stake, participants report exerting higher listening effort (Carolan et al., 2021) or show physiological signs of doing so, as measured by pupillometry and cardiovascular reactivity (Koelewijn et al., 2018, 2021; Richter, 2016)—without, however, experiencing corresponding performance benefits. Plain and colleagues (2021), on the contrary, report small behavioural improvements associated with monetary incentives with no corresponding physiological evidence for the effect of reward on listening effort. A positive effect of reward on *both* listening effort and recognition of perceptually difficult speech was attested in only one study (Zhang et al., 2019)—but these effects pertain to spectrally-intact and unmasked time-compressed speech. At the same, there is abundant evidence that monetary reward improves the key aspects of performance in a variety of cognitive tasks (Botvinick & Braver, 2015; Krebs & Woldorff, 2017; Notebaert & Braem, 2016; Pessoa, 2015), including the accuracy of perceptual decision making with degraded *visual* stimuli (Blank et al., 2013; Engelmann et al., 2009). It is therefore of interest whether similar reward effects can be obtained on a perceptual decision task involving degraded *spoken* stimuli—that is, whether the increased engagement and listening effort associated with reward can translate into a more accurate perception of degraded speech.

The present study uses monetary incentives to manipulate listeners' attentional engagement on a rewarded version of the same/different task—the classic induced-misperception set-up used in previous studies with noise-vocoded speech (Blank et al., 2018; Sohoglu et al., 2014). To induce misperception on a subset of trials, we provide prior expectations by presenting matching, mismatching, or partially mismatching written text just before the degraded speech. We further manipulate the relative value of accurate perception in a block- and trial-by-trial basis. A fixed monetary bonus is offered for correct response on some—but not all—trials in a reward block. A baseline block, performed without any knowledge of incentives and before the reward block, serves as a control measure of intrinsic motivation, unaffected by external reward manipulations. Incentive trials in the reward block were expected to have higher utility than non-incentive and baseline trials. We, therefore, predicted that these trials would strongly engage attentional resources and cognitive control, enhancing listeners' sensitivity to mismatching prior information. In the case of partial mismatch trials, this was expected to translate into a lower rate of misperception on incentive trials—relative to non-incentive and baseline trials. But since behavioural measures in speech perception tasks do not always accurately reflect the degree of attentional engagement and exerted listening effort, and since these two tend to dynamically change even over the course of a single trial, we chose to use pupillometry as an additional, time-sensitive measure of attention during listening.

1.3 Pupillometry as an index of attentional engagement

Previous studies have linked elevated listening effort to increased pupil dilation (Alhanbali et al., 2020; Miles et al., 2017; Winn, 2016; Winn et al., 2015; Winn & Teece, 2021; Zekveld et al., 2010, 2018; Zhao et al., 2019). Under controlled luminance, pupil dilates in response to growing task demands, reflecting changes in cognitive load and task engagement mediated by locus coeruleus (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011). The locus coeruleus-norepinephrine (LC-NE) system controls changes in the attentional state via two modes of function (Aston-Jones & Cohen, 2005). Phasic LC-NE activity facilitates task-relevant behaviours and optimizes within-task performance through adaptive gain. Tonic LC-NE activity optimizes performance across tasks by increasing neuronal sensitivity to task-irrelevant stimuli. The neuromodulatory activity of the LC-NE system is receptive to the ongoing evaluation of task utility: phasic LC-NE activity supports task performance only while the task utility remains reasonably high. Once the utility of the current task drops below a certain threshold, the LC-NE system withdraws adaptive gain support and shifts into the tonic mode of function, facilitating other, more “exploratory” forms of behaviour (Aston-Jones & Cohen, 2005). Pupil dilation dynamics closely track these changes in LC-NE activity (Gilzenrat et al., 2010), serving as a physiological marker of exerted mental effort across a range of cognitive tasks (van der Wel & van Steenbergen, 2018). A stimulus-driven peak pupil dilation corresponds to a phasic NE release and reflects focused attention and adaptive control (Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011). Elevated baseline diameter coupled with a suppressed peak dilation response reflects high tonic NE, associated with scanning attention and mind wandering. As such, pupil size is considered a reliable measure of effort and attentional engagement during listening—more reliable than accuracy and intelligibility scores (Winn et al., 2015).

Pupil dilation predictably follows the same inverse U-shaped response function as other indices of adaptive control processes, including neural activity in cingulo-opercular and frontoparietal regions and behavioural performance measures (Zekveld et al., 2018). On listening-in-noise tasks, mean pupil dilation response and peak dilation (i.e., phasic response) increase with decreasing speech intelligibility—up until the point when speech recognition becomes practically impossible (Miles et al., 2017; Zekveld et al., 2010; Winn et al., 2015; Winn, 2016). Very low speech intelligibility is conversely associated with a considerably smaller phasic response relative to baseline (i.e., the resting-state pupil diameter)—which is frequently interpreted as an index of task withdrawal (Zekveld et al., 2010). Because the role of cognitive control is to support optimal performance on challenging tasks, pupil size often reflects not only listening demand and listening effort but also perceptual outcomes. Stronger peak dilation and larger pre-stimulus pupil size have been linked to higher accuracy on speech recognition tasks (Alhanbali et al., 2020; Zekveld et al., 2010). In contrast, smaller pre-stimulus pupil size and elevated baseline diameter are associated with poor performance, listening fatigue, and subsequent task disengagement (Alhandbali et al., 2020; Gilzenrat et al., 2010; Zekveld et al., 2010). Like cognitive control, pupillary response is affected by changes in motivation and task utility. High monetary rewards drive larger stimulus-evoked dilations than low monetary rewards, reflecting stronger adaptive gain support for higher-utility tasks (Knapen et al., 2016). Pupillometry has been shown to reflect even momentary fluctuations of attention, and as such represents a reliable, objective, and, more importantly, time-sensitive measure of attentional engagement on listening trials (Kang & Wheatley, 2015; Wierda et al., 2012; Zénon, 2017; Zhao et al., 2019)—the perfect tool for the present study.

We, therefore, used pupillometry as an index of moment-to-moment attentional engagement during the rewarded version of the same/different task—a perceptual decision task with degraded spoken words that were preceded by matching, mismatching, or partially mismatching written text. We recorded listeners' pupil size as they approached each trial, thus tracking perceptual processing at different levels of motivation—i.e., being intrinsically motivated throughout the baseline block, extrinsically motivated on incentive trials within the reward block, and mildly demotivated on non-incentive trials within the same block. This allowed us to investigate how monetary incentives affected listeners' attentional engagement during the trial and how pupil dilation trajectories reflected perceptual outcomes on partial mismatch trials (i.e., misperception vs veridical perception). We hypothesized that incentive trials would exhibit stronger phasic dilation relative to both baseline and non-incentive trials, regardless of the perceptual outcome. Misperceived trials were expected to be preceded by smaller pre-stimulus dilation, than correctly perceived trials. We had no specific hypotheses concerning the reward-by-accuracy interaction, so we used generalized additive mixed models (GAMM) to map the time course of pupil dilation for misperceived and correctly perceived partial mismatch trials across the three incentive conditions (baseline, incentive, non-incentive) and compare these effects directly.

Chapter 2: Methods

2.1 Design

To investigate the effect of attention and prediction on (mis)perception of degraded speech, perceptual and pupillary responses were acquired in an experiment using a mixed block/event 3x3 within-subject design (Chiew & Braver, 2013); see Figure 1 for an illustration of the experimental design. Attention to stimuli pairs was manipulated at both block and trial levels by setting three conditions: baseline, incentive, and non-incentive. Participants performed separate baseline and reward blocks, in this fixed order. The baseline block was performed without any knowledge of incentives. Within the reward block, incentive trials were randomly intermixed with non-incentive trials. This design allows us to examine trial-based effects of reward (by contrasting incentive with non-incentive trials within the reward block) while controlling for block-based changes in attention and motivation (by contrasting non-incentive trials against the baseline trials). The main reason for using these contrasts was that previous research demonstrated that these block- and trial-level incentive effects are accompanied by a rather different pupil dilation profiles and may correspond to different underlying mechanisms (Chiew & Braver, 2013; Jimura et al., 2010).

Prior expectations were induced by presenting written words before degraded spoken words. The validity of predictions was manipulated in three conditions: (1) matching written text (*cap–cap*), (2) mismatching written text (*cap–win*), or (3) partially mismatching written text (onset mismatch: *cap–map*, and offset mismatch: *cap–cat*). Thus, global (experiment-wise) prior validity was set to 0.25—a prior matched the following degraded word only on 25% of trials. Each condition involved 32 different word pairs that were repeated three times throughout the experiment—once within each condition (baseline, incentive, non-incentive). Partial mismatch trials, which comprised half of the dataset, were the main target of analyses.

Behavioural responses were collected in a 2AFC same/different task. In each trial, listeners indicated whether they thought that the degraded word matched the presented written word by pressing one of two buttons (s = "same", d = "different"). Pupil size was recorded while participants performed the task to investigate how reward affected perceptual processing.

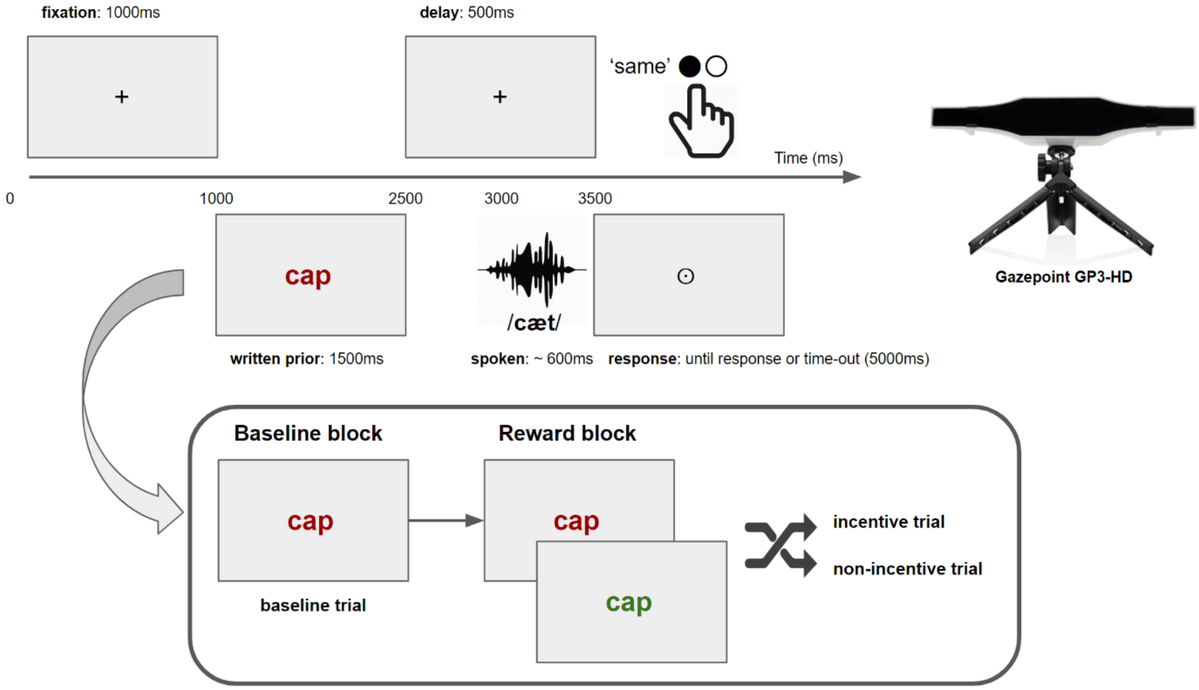


Figure 1. Experimental design. Listeners heard degraded spoken words preceded by matching, mismatching, or partially mismatching written text and responded whether the two words were the same or different. There were two blocks: a baseline block performed without any knowledge of incentives and a reward block where incentive and non-incentive trials were randomly intermixed. In the reward block, the color of the written word reflected the trial type (incentive vs non-incentive).

2.2 Participants

Fifty-one native English speakers (40 females, 9 males, 2 non-binary; mean age 22.14 ± 4.49) were recruited using email solicitation. All participants provided informed consent and reported having normal or corrected-to-normal vision and no history of hearing, linguistic or cognitive impairments. Participants were paid at the fixed rate of C$15/hour, plus a C$2.8– C$9

(M = $6.34 ± 1.20) bonus based on their task performance. Ethics approval was provided by the Research Ethics Board of York University (Certificate #: STU 2022-014).

Due to poor tracking quality during testing, data from four participants were discarded. Subsequent analyses were performed using data from 47 participants (37 females, 8 males, 2 non-binary; mean age 22.11 ± 4.44).

2.3 Stimuli

The stimulus set consisted of 32 monosyllabic English words presented in written and spoken format. The words were recorded by a native speaker of North American English at 16 bit with a sampling rate of 44.1 kHz. The duration of spoken words ranged 609–829 ms (M = 703.6 ms, SD = 50.7 ms). Recorded stimuli were intensity-scaled (60 dB) and then noise-vocoded in Praat (Boersma & Weenink, 2021), using a modified version of a script originally written by Darwin (n.d.) and following the previously described protocol (Davis et al., 2005). The words were first filtered into six logarithmically spaced frequency bands spanning between 50 and 8000 Hz. Contiguous band-pass filters were constructed in the frequency domain: passbands were 3 dB down at 50, 229, 558, 1161, 2265, 4290 and 8000 Hz with a roll-off of 22 dB per octave. The amplitude envelope from each frequency band was extracted using a standard Praat algorithm (squaring intensity values and convolving with a 64-ms Kaizer-20 window, removing pitch-synchronous oscillations above 50 Hz). The resulting envelope was then applied to band-pass filtered noise in the same frequency ranges. Finally, the resulting bands of modulated noise were recombined to produce the degraded word. These parameters were chosen based on the previous work that has shown high accuracy for match and mismatch conditions and high response variability on partial mismatch trials (Blank et al., 2018; Sohoglu et al., 2014).

In line with Blank et al (2018), the 32 words formed two sets of 16 words, each set containing four quadruplets of words deviating in either onset or offset sounds (see Figure 2A for the illustration). Written and spoken words were combined in three conditions: (1) 32 match pairs (e.g., *cap–cap*), (2) 32 mismatch pairs (*cap–win*), and (3) 64 partial mismatch pairs (32 onset mismatch: *cap–map*, and 32 offset mismatch: *cap–cat*; see Table 1 for a full list of partial mismatch pairs). Each written-spoken pair was repeated three times throughout the experiment—once at each level of reward (baseline, incentive, non-incentive). Each word occurred in its written or spoken form with equal probability, 24 times in total (twelve times as a prior, twelve times as a spoken degraded word).

Table 1. Full list of partial mismatch pairs.

			onset mismatch				**offset mismatch**			
word pair	**vowel number**	**quadruple**	**written word**	**spoken word**	**deviating sound**	**common sound**	**written word**	**spoken word**	**deviating sound**	**common sound**
1	1	1	cap	pap	k/p	æp	cap	cat	p/t	kæ
2	1	1	cat	pat	k/p	æt	cat	cap	t/p	kæ
3	1	1	can	pan	k/p	æn	can	cam	n/m	kæ
4	1	1	cam	Pam	k/p	æm	cam	can	m/n	kæ
5	1	1	pap	cap	p/k	æp	pap	pat	p/t	pæ
6	1	1	pat	cat	p/k	æt	pat	pap	t/p	pæ
7	1	1	pan	can	p/k	æn	pan	Pam	n/m	pæ
8	1	1	Pam	cam	p/k	æm	Pam	pan	m/n	pæ
9	1	2	bap	map	b/m	æp	bap	bat	p/t	bæ
10	1	2	bat	mat	b/m	æt	bat	bap	t/p	bæ
11	1	2	ban	man	b/m	æn	ban	bam	n/m	bæ
12	1	2	bam	mam	b/m	æm	bam	ban	m/n	bæ
13	1	2	map	bap	m/b	æp	map	mat	p/t	mæ
14	1	2	mat	bat	m/b	æt	mat	map	t/p	mæ
15	1	2	man	ban	m/b	æn	man	mam	n/m	mæ
16	1	2	mam	bam	m/b	æm	mam	man	m/n	mæ
17	2	3	tip	kip	t/k	ɪp	tip	tit	p/t	tɪ
18	2	3	tit	kit	t/k	ɪt	tit	tip	t/p	tɪ
19	2	3	tin	kin	t/k	ɪn	tin	Tim	n/m	tɪ
20	2	3	Tim	Kim	t/k	ɪm	Tim	tin	m/n	tɪ
21	2	3	kip	tip	k/t	ɪp	kip	kit	p/t	kɪ
22	2	3	kit	tit	k/t	ɪt	kit	kip	t/p	kɪ
23	2	3	kin	tin	k/t	ɪn	kin	Kim	n/m	kɪ
24	2	3	Kim	Tim	k/t	ɪm	Kim	kin	m/n	kɪ
25	2	4	whip	lip	w/l	ɪp	whip	wit	p/t	wɪ
26	2	4	wit	lit	w/l	ɪt	wit	whip	t/p	wɪ

27	2	4	win	Lynn	w/l	ɪn	win	whim	n/m	wɪ
28	2	4	whim	limn	w/l	ɪm	whim	win	m/n	wɪ
29	2	4	lip	whip	l/w	ɪp	lip	lit	p/t	lɪ
30	2	4	lit	wit	l/w	ɪt	lit	lip	t/p	lɪ
31	2	4	Lynn	win	l/w	ɪn	Lynn	limn	n/m	lɪ
32	2	4	limn	whim	l/w	ɪm	limn	Lynn	m/n	lɪ

2.4 Procedure

Stimulus presentation and data collection were controlled using a custom Python script built upon PsychoPy (Peirce et al., 2019) and PyGaze (Dalmaijer et al., 2014) libraries. Participants were seated in front of a 23.8” monitor in a moderately-lit laboratory of approximately 250 lx, with their head stabilized with a chinrest. Monitor height was adjusted such that participants’ eyes were positioned halfway on the screen when looking straight ahead (Ooms et al., 2015). Visual stimuli were presented at the center of the screen, subtending approximately 3 degrees of visual angle. Auditory stimuli were presented binaurally through over-ear Sennheiser HD 400S headphones, at a comfortable volume.

Participants completed two blocks of the same/different task, presented in a fixed order: a baseline block followed by a reward block. The baseline block consisted of 128 trials. This block was performed without any knowledge of incentives. The reward block consisted of 256 trials (128 incentive trials and 128 non-incentive trials, randomized). In this block, participants could either earn or lose a $0.1 bonus by giving, respectively, a correct and incorrect response on a subset of trials (specified by the color of the written word: red or green). Each set of 128 trials (baseline, incentive, non-incentive) consisted of 32 match trials, 32 mismatch trials, and 64 partial mismatch trials, randomized.

Each trial began with a central fixation cross (1000 ms), followed by a visually presented “prior”, or written word (1500 ms). In the reward block, its color (red or green) served as an incentive cue signaling incentive and non-incentive trials. In the baseline block, the text color was kept constant (i.e., all words were either red or green). The mapping between text color and the three levels of incentive (baseline, incentive, non-incentive) varied randomly across participants. The duration of the presentation of the prior—1500ms—was chosen to ensure that listeners had sufficient time to process both the meaning and the incentive information encoded in written text (Chiew & Braver, 2016). Previous research demonstrated that the effects of prior knowledge remain robust regardless of stimulus onset asynchrony, as long as the written text is presented before the degraded speech (Sohoglu et al., 2014). Therefore, following a 500 ms delay, participants heard a degraded spoken word and were prompted to respond whether the two words were same or different. Listeners responded by pressing one of two buttons on a keyboard (“s” – same; “d” – different) in the next 5000 ms window. Trials were separated by a 1000 ms inter-trial interval.

To minimize any feedback-related changes in motivation and motivation-induced learning effects (Notebaert & Braem, 2016), participants were not given trial-by-trial feedback. Instead, feedback was presented every 32 trials (every 2-3 minutes). In the baseline block, the feedback message read “Set over. Starting the next set”. In the reward block, the feedback message read “Your bonus so far is $X”, indicating the amount earned by a participant by that point.

Subjective intelligibility of noise-vocoded speech is highly dependent on experience (Davis et al., 2005), so prior to the experimental blocks, all participants completed a practice session that was identical—in all respects—to the baseline block. The purpose of this session was to familiarize listeners with noise-vocoded speech, allow initial perceptual learning to take place, and ensure asymptotic performance. A short 8-trial training session also preceded the reward block, allowing participants to practice the association between the color of written word and trial type (i.e., incentive vs non-incentive).

The entire experiment took about 50 minutes to complete. Participants were given self-timed breaks between the blocks.

2.5 Pupillometry data collection and preprocessing

Pupil size was continuously recorded from each eye using a Gazepoint GP3-HD infrared eye-tracker at a sampling rate of 150 Hz. The eye-tracker was mounted directly under the screen, situated 60 cm away from the participant. Screen brightness was adjusted to intermediate levels (50%) to approximately match the luminance of the room and avoid discomfort glare. All text colors used in the experiment (green, red, gray) were matched for relative luminance (Y = 0.293 ± 0), perceived lightness (L* = 61.1 ± .027) and perceived contrast against a light-beige background (-0.569 ± .001)[2].

At the beginning of each experimental run, the eye-tracker was calibrated using a nine-point calibration procedure. Calibration was accepted only when average accuracy in vertical and horizontal dimensions for both eyes was below 40 px (about 1 degree of visual angle), with all points valid.

Before analysis, pupil data were preprocessed in R (R Core Team, 2020). Samples tagged as invalid by the eyetracker were removed. Trials with more than 30% of invalid data points were excluded from the analysis (n = 671, 2.85%). Since Gazepoint GP3-HD eyetracker does not

[2] Relative luminance (Y) is a linear measure of light. It is based on spectral sensitivity of human vision, so it reflects how human eye perceives different wavelengths of light (color). However, luminance is not adjusted for the non-linear perception of lightness. Y ranges from 0 to 1, with 0 being black and 1 being white. Perceptual lightness (L*) reflects visual perception of luminance and can be approximated from Y. Contrast refers to the difference between two Y or L* values, for foreground and background colors respectively (*Myndex^{TM} Web Help - Luminance Contrast and Perception*, n.d.). Here, perceived contrast was calculated.

automatically detect blinks, a simple velocity filter (Mathôt, 2013) was used for deblinking. The remaining artifacts were removed using the approach described in Kret & Sjak-Shie (2019). Dilation speed outliers—samples at which dilation speed exceeded a threshold based on median absolute deviation (MAD) from dilation speed series—were detected and removed. Clusters of samples that strongly deviated from the signal trend line, generated by interpolating and smoothing filtered data, were identified and removed in a similar fashion. Blinks (continuous stretches of missing data lasting 75–500ms) were extended 25ms forward and backward. Finally, a sparsity filter was applied to reject "islands"—short (<50ms) clusters of temporally isolated samples, likely attributed to noise, that remained after previous filtering steps. Visual inspection followed automatic artifact rejection.

Because right and left pupil sizes are highly correlated (Jackson & Sirois, 2009), the analyses were based on mean pupil data (Kret & Sjak-Shie, 2019). To generate these times series, a dynamic offset between the left and right pupil diameter was calculated when samples were available for both eyes. To estimate mean pupil diameter in the presence of missing samples, this offset was interpolated to the time points that only had data from one eye. Next, to reduce autocorrelation in the subsequent GAM analysis (van Rij et al., 2019), these averaged pupil time series were filtered using a 5-th order zero-phase low-pass Butterworth filter with a cut-off frequency of 10 Hz and downsampled to 50 Hz by taking the mean per time-bin in. In the process, all stretches of missing data shorter than 400ms were interpolated. Downsampled pupil data were baseline-corrected, using the subtractive method (Mathôt et al., 2018). The baseline was calculated per trial as the median pupil size during the 100 ms following the onset of the written word. Since the latency of the fastest pupil response is about 200 ms (Mathot, 2018), pupil size in this period was not affected by any experimental manipulation, while its relatively short duration reduced the likelihood of artifacts and pupil-size fluctuations. If baseline correction could not be performed due to missing data (e.g., because of a blink during the baseline period), the trial was removed. Trials with extreme baseline pupil sizes (with the z-scored baseline being larger than 3 or smaller than -3) were likewise removed (Mathôt & Vilotijević, 2022). Finally, the velocity filter was reapplied to remove the remaining recording artifacts. A total of 1078 trials (5.97%) were excluded from the analysis.

The data were aligned to the onset of the written word (2000 ms before the presentation of degraded speech). The pupil dilation in the period of 4000 ms from the onset of the prior (i.e., 2000 ms before and 2000ms after the onset of the degraded word) was analyzed.

2.6 Analyses

All analyses were performed in R (version 4.1.3: R Core Team). To assess whether perception was influenced by reward, behavioural responses on partial mismatch trials were analyzed using a mixed logistic model. Pupillometry time series were analyzed using generalized additive mixed models.

2.6.1 Acoustic similarity analysis

Acoustic similarity between degraded spoken words was computed following the previously described protocol (Billig et al., 2013; Blank et al., 2018). First, a gammatone-based spectrogram-like time-frequency matrix was computed for each degraded word, to approximate the frequency analysis performed by the human ear. Then, for each pair of tokens, a spectral similarity matrix was generated by comparing gammatone spectral profiles of all time slices. We chose to compute the spectral similarity between six-channel vocoded versions of written and spoken pairs because it was shown to have a higher correlation with perceptual outcomes than the similarity between noise-vocoded and clear words (Blank et al., 2018). Dynamic time warping was used to compute the maximum similarity path through each similarity matrix. Finally, the summed similarity values along this path were computed and normalized. Resulting (dis)similarity scores ranged from 0 (for the two most similar sound files) to 1 (for the two most dissimilar sound files). Acoustic similarity analysis was performed in MATLAB using existing gammatone spectral analysis and dynamic time warp functions supplied by Ellis (2003, 2009). Figure 2B illustrates the computed spectro-temporal similarity between noise-vocoded versions of tokens in each written/spoken pair.

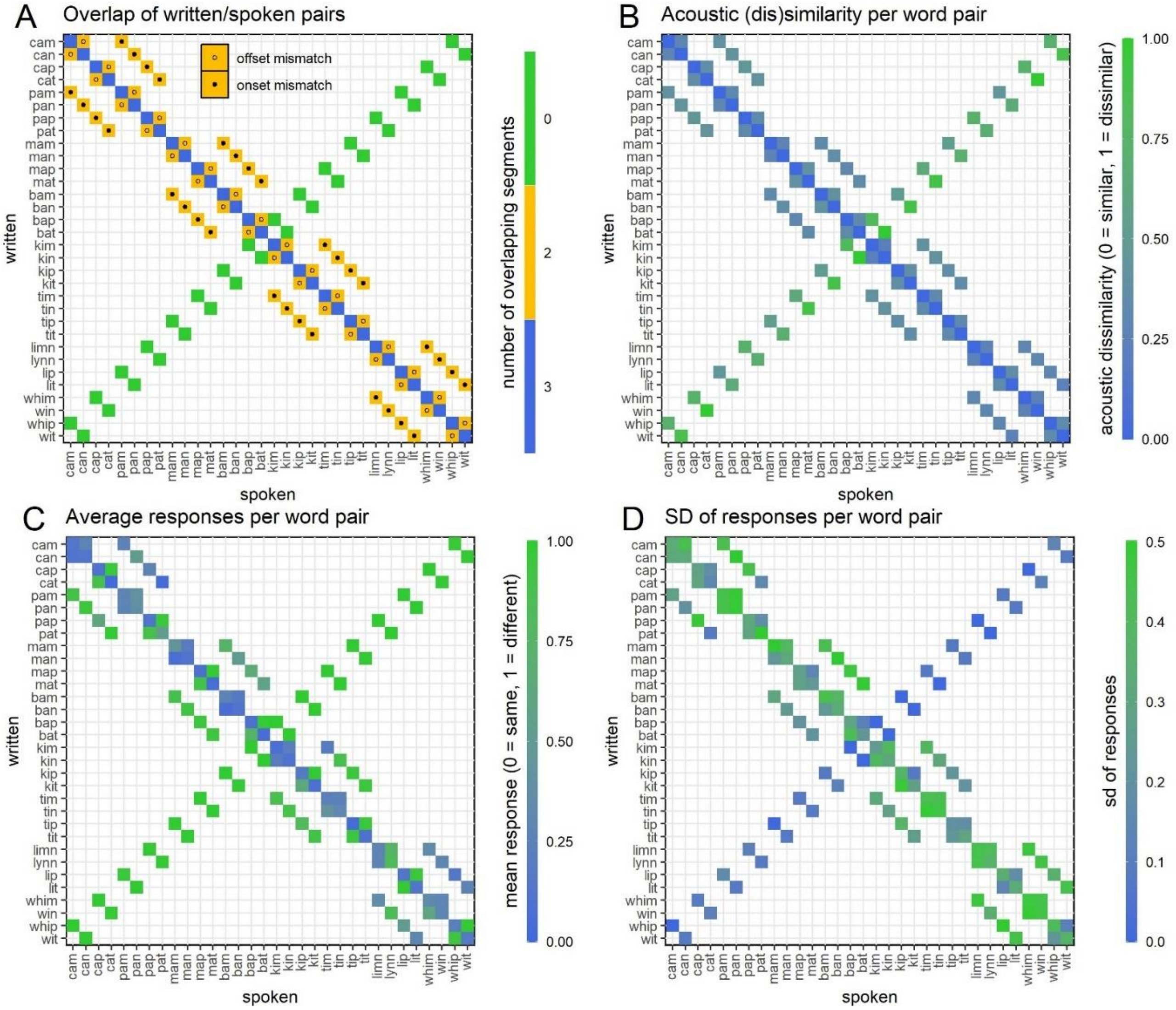


Figure 2. Stimulus similarity and behavioural confusion matrices. (A) Word similarity matrix. 32 written and spoken words were combined in three different conditions:32 match pairs (3 overlapping segments), 64 partial mismatch pairs (2 overlapping segments: in onset and offset respectively), 32 total mismatch pairs (no overlapping segments). (B) Acoustic similarity between

noise-vocoded versions of written and spoken words. (C) Mean behavioural responses. Mismatching pairs were more often reported as "different" (green), matching pairs—as "same" (blue), while responses to partial mismatch pairs were a mix of "same" and "different". (D) SD of responses. Responses to mismatching and matching pairs were more consistent (blue), responses to partially mismatching pairs—more variable (green).

2.6.2 Logistic mixed-effects regression analysis of the behavioral data

First, we used a logistic mixed model in lme4 (Bates et al., 2015) to analyze all behavioural responses. This model aimed to (1) estimate how (mis)informative prior knowledge affected perception of noise-vocoded words and (2) investigate whether incentives affected perceptual outcomes of listening beyond the prediction effects. To do so, we modeled trial-by-trial accuracy as a function of prior (matching, mismatching, partially mismatching), reward condition (baseline, incentive, non-incentive), and their interaction. We further included random intercepts for item and participant to account for the lack of independence on the participant- and item levels.

Perceptual outcomes of partial mismatch trials—the main target of our analyses—were analyzed using another mixed logistic model. This model was fitted using the following lme4 specification:

```
correct ~ reward + sim.z + (reward  | id) + (1 | pair)
```

Here, the dependent variable was accuracy on a single trial (0: incorrect/misperception, 1: correct/veridical perception). The model included the fixed effects of reward and acoustic similarity. The 3-level factor of reward (baseline vs incentive vs non-incentive) was coded using repeated contrasts. This allowed us to test both the trial-based effect of reward (by comparing baseline and non-incentive trials) and the block-based effect of reward (by comparing incentive and non-incentive trials). Acoustic similarity, a continuous predictor, was standardized: centered at 0 and divided by 2 standard deviations (Gelman, 2008). Standardization placed this term on the same scale as binary input variables, allowing for a more intuitive interpretation of model coefficients. To test whether an interaction term (between reward and acoustic similarity) was necessary, we fitted a model with this interaction term and performed model comparison using a likelihood ratio test. The test indicated that the interaction was not necessary ($\chi2 = 0.16$, $p = 0.92$), so this term was not included in the final model.

To capture the multilevel structure of the data (repeated measures for both participants and items), the model included the maximal random effect structure justified by the data: random intercepts for participants and items, as well as a random slope for reward within participants. This random effect structure captured intrinsic variability in perception among participants, individual differences in reward sensitivity, as well as varying perceptual difficulty of stimuli. For items, although the same 32 words appeared as both written primes and spoken words (see Figure 2A), it was decided to collapse them into a single 64-level random factor coding a written/spoken "pair", instead of fitting separate random effects for written priors and spoken words. Timed-out trials were removed from the analysis (1.4% data loss).

Since regression allows us to test the influence of potential confounding variables, we fit another model including two of such nuisance factors: time-within-experiment and reaction time. Given that the order of baseline and reward blocks was fixed, the effects of time-within-experiment are particularly important to consider as a potential confound. Previous work (Davis et al., 2005) has shown robust perceptual learning effects after the first few minutes of exposure to noise-vocoded speech. Perceptual learning tends to follow a power-law curve with most learning happening rather quickly, and we included a practice session to stabilize performance and mitigate these effects. However, it is still possible that slower learning-related changes affected performance in the rest of the experiment, since the dataset was rather small, and listeners heard (and saw) each word multiple times. In a similar vein, reaction time is an important source of concern because reward manipulations in other experiments were shown to speed up reaction times, to the detriment of accuracy (Bogacz et al., 2006; Dambacher et al., 2011; Drugowitsch et al., 2015)—the effect known as speed-accuracy trade-off (SAT). Therefore, we fit another model that included these potential sources of variability. As before, both continuous variables were standardized: centered and divided by 2 standard deviations. RT was standardized within participant.

The models were fit using maximum likelihood (Laplace Approximation). Likelihood ratio tests were used to test statistical significance of predictors, as implemented in lme4 (Bates et al., 2015). Since Bayesian models offer more robust estimation, especially in the context of clustered binary data (Fong et al., 2010), confidence intervals and point estimates of the final model coefficients were obtained by fitting analogous Bayesian models in Stan and R using brms package (Bürkner, 2017). To improve model convergence and guard against overfitting, we specified independent Cauchy priors on all logistic regression coefficients, each centered at 0 and with scale parameter 10 for the intercept and 2.5 for all other coefficients (Gelman et al., 2008). These prior distributions can be directly interpreted as a constraint on the coefficients: in combination with standardization, they imply that the absolute difference in logit probability when moving from one SD below the mean to one SD above the mean, should be less than 5 for any given input variable. In logistic regression, a change of 5 on the logit scale corresponds to 48% increase in log odds—a shift in probability from 0.02 to 0.5 or from 0.5 to 0.98. Since it is highly unlike that any single term changes the probability of the outcome from 0.02 to 0.98, we reflect this in the model by assigning low probabilities to changes of 10 and higher on the logit scale. We further specified half-Student-t priors with 3 degrees of freedom, center 0, and scale 2.5 on all variance parameters. While still being weakly informative, this prior has lighter tails than the recommended half-Cauchy prior (Gelman, 2006; Polson & Scott, 2012). This, in turn, leads to better convergence and more robust estimates in the case of logistic regression, where the likelihood is highly sensitive to large values of the underlying linear predictor. The model was sampled using 4 chains, 8000 iterations each, with 2000 iterations used for warmup—returning a total of 24000 post-warmup samples. Samples were drawn using the No-U-Turn-Sampler (NUTS: Hoffman & Gelman, 2014). Visual model diagnostics was performed using bayesplot package (version 1.9.0: Gabry et al., 2018). All R-hat values were < 1.005, and all ratios of the effective sample size to the total sample size were above 0.1, with most above 0.5, indicating good convergence.

2.6.3 GAMM analysis of the pupil data

Pupil time series were analyzed with a generalized additive mixed model using mgcv R package (version 1.8.39: Wood et al., 2013) and visualized using itsadug R package (version 2.4: Rij et al., 2022). Generalized additive models (Wood, 2017) are an extension of linear regression in that they assess the relationship between a dependent variable and a number of available predictors. But while linear regression assumes this relation to be linear, GAMs allow us to model it as a smooth function—a continuous, potentially wiggly, line that changes over time to fit the (non-linear) pattern of the data. These models do not require pre-defined non-linear function specification—instead, smooths are approximated from the data as a weighted sum of multiple base functions, such as cubic and thin plate regression splines. GAMs use penalized regression methods (i.e., they penalize wiggliness—model complexity—in favor of simpler (non)linear trajectories) to obtain the maximum likelihood estimation of the smooths. Smoothing parameters of multiple predictors are estimated using cross-validation, which allows GAMs to avoid both overgeneralization and overfitting. Generalized additive *mixed* models, in particular, are well suited for analyzing the time-course of pupillometric data, because of their flexibility and ability to detect non-linear effects, while simultaneously accounting for variation in both participants and items (van Rij et al., 2019).

Here, to examine the possibly (non-linear) effects of reward and perceptual outcome on the pupil dilation trajectory over time, we looked at the time-locked baseline-corrected pupil size in the time range of -2000 to 2000 ms before/after audio onset. Note that our dependent variable (the change in pupil size relative to baseline) is expressed in pixels as reported by GazePoint; currently, there is no straightforward way to reliably convert these values to the metric scale.

The model was fitted using the bam() function from the mgcv package (version 1.8-39) with fREML estimation and discretized covariates for faster computation (i.e., with arguments method="fREML" and discrete = TRUE). The model was specified in the following way:

```
PD ~ s(timebin, by = correct, k = 20) + correct
+ s(timebin, by = Is1B)
+ s(timebin, by = Is1I)
+ s(timebin, by = Is0B)
+ s(timebin, by = Is0I)
+ s(timebin, id, by = correct, bs = "fs", m = 1)
+ s(timebin, id, by = Is1B, bs = "fs", m = 1)
+ s(timebin, id, by = Is1I, bs = "fs", m = 1)
+ s(timebin, id, by = Is0B, bs = "fs", m = 1)
+ s(timebin, id, by = Is0I, bs = "fs", m = 1)
+ s(timebin, pair, by = correct, bs = "fs", m = 1)
+ s(timebin, pair, by = Is1B, bs = "fs", m = 1)
+ s(timebin, pair, by = Is1I, bs = "fs", m = 1)
+ s(timebin, pair, by = Is0B, bs = "fs", m = 1)
+ s(timebin, pair, by = Is0I, bs = "fs", m = 1)
```

This specification indicates that we are modeling non-linear effects of reward and subsequent perceptual outcome on pupil dilation within the trial. Essentially, we are fitting these effects for three types of reward trials and for both misperceived and correctly perceived word pairs. But instead of obtaining separate dilation trajectories for each of these six combinations, we model both trial- and block effects of reward directly, using binary difference smooths

(Sóskuthy, 2017; Wieling, 2018). Binary difference smooths integrate the constant and non-linear difference between two categories into a single term and evaluate whether this term is necessary (i.e., whether the difference wave between two categories is significantly different from 0). Modeling target differences with binary smooths produces the highest power compared to other methods of significance testing (such as model comparison), without increasing computational cost or Type I error rate (Sóskuthy, 2021). To implement by-factor binary smooths, we converted *Reward* into a set of binary predictors corresponding to each of our four differences of interest: Is1B (1: correct baseline trials; 0 otherwise), Is1I (1: correct incentive trials; 0 otherwise), Is0B (1: misperceived baseline trials; 0 otherwise) and Is0I (1: misperceived incentive trials; 0 otherwise). The first term in the model specification, then, fits two *reference* curves corresponding to pupil trajectories for misperceived and correctly perceived non-incentive trials. The second term is a constant; it models the parametric difference between correct and incorrect non-incentive trials. The next four terms correspond to *difference* curves: they directly model the block-level and trial-level effects of incentive that we are interested in. Essentially, *s(timebin, by = Is1B)* estimates the difference between *correct* baseline and non-incentive trials, *s(timebin, by = Is1I)*—between *correct* incentive and non-incentive trials, while *s(timebin, by = Is0B)* and *s(timebin, by = Is0I)* model respective differences for *incorrect* trials.

To account for the hierarchical structure of the data, the model included random smooths for participants and items (the last ten terms in the model specification). Random smooths are analogous to random intercept and slope adjustment in linear mixed effect models but are more flexible because they adjust the entire shape of a (non-)linear regression line (which may or may not involve adjustment of the intercept and/or slope). The ten random smooths fitted here estimate the individual variability in the pupil dilation trends per participant and word pair (separately for the reference curve and four difference curves, due to by-factor specification). Such a “random reference/difference” approach is optimal when modeling within-unit random effects, such as those attested here. It offers higher power under the same nominal Type I error rate compared to other random effect structures, such as item*effect smooths and item-by-effect smooths (Sóskuthy, 2021). Note that Sóskuthy (2021) referred to *ordered factor* random difference smooths instead of binary ones, although these are essentially the same, given that the random-effect smooths—regardless of whether they are binary or ordered factor—are not centered (i.e., they always include intercept: Sóskuthy, 2017; Wieling, 2021a).

Both fixed and random smooth terms were constructed using thin plate regression smooths. The maximum number of basis functions was set to 19 for the fixed reference smooth and to 9 for all binary difference smooths (i.e., k = 20 and k = 10 respectively). Model checking with gam.check() indicated that these numbers were sufficient. To correct for autocorrelation in residuals, which is rather extreme in pupillometric signal (van Rij et al., 2019), the fitted model included an AR1 autoregressive error model. This model directly estimates the effect of an immediately preceding data point on the current one. The correlation parameter (rho) for an AR1 model was estimated using the residual autocorrelation at lag 1 of an identical model fitted to the same data without an AR1 component (Sóskuthy, 2021). Finally, since the distribution of the measured pupil signal was non-normal, resulting in non-normal residuals, the model was re-fitted using the scaled t-distribution, better suited for heavy-tailed response variables (family = ‘scat’). As a result of this change, the residuals substantially improved. Diagnostic plots for the model are provided below (Figure 3).

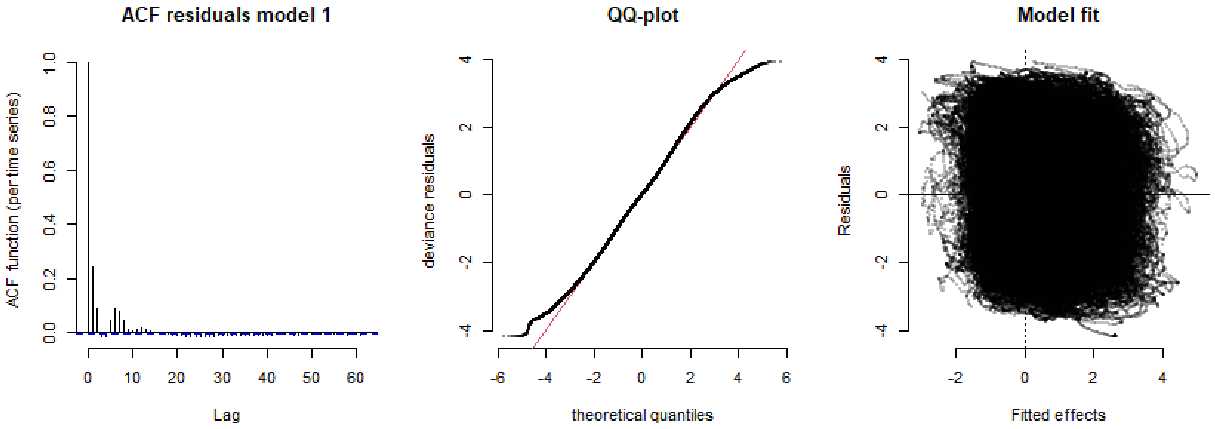


Figure 3. Diagnostic plots for the GAMM model. ACF = autocorrelation function ; QQ = quantile-quantile.

To assess whether the differences between reward conditions for correct and incorrect trials were constant (i.e., a difference in height only) or non-linear (i.e., a difference in the shape of the curves), we refit the model above using ordered factors (Sóskuthy, 2017; Wieling, 2018, 2021b). Every binary smooth (Is1B, Is1I, Is0B, and Is0I) was therefore converted into a corresponding ordered factor (Is1B.o, Is1I.o, Is0B.o, and Is0I.o) and re-entered in the model as a parametric effect and an ordered factor difference smooth. Essentially, in this second model, the four binary difference factor smooths were split into four parametric terms and four ordered factor smooths. The random-effects specification is essentially the same as that of the previous model: ordered factors were included as random smooths to account for individual differences in the pupil trajectories for participants and word pairs. As noted earlier, in this case, it does not matter whether random effects are fitted using ordered factor smooths or binary difference smooths since both are not centered. The full model specification is provided below; all differences from the binary curve model are in bold:

```
PD ~ s(timebin, by = correct, k = 20) + correct
+ s(timebin, by = Is1B.o) + Is1B.o
+  s(timebin, by = Is1I.o) + Is1I.o
+ s(timebin, by = Is0B.o) + Is0B.o
+ s(timebin, by = Is0I.o) + Is0I.o
+ s(timebin, id, by = correct, bs = "fs", m = 1)
+ s(timebin, id, by = Is1B.o, bs = "fs", m = 1)
+  s(timebin, id, by = Is1I.o, bs = "fs", m = 1)
+ s(timebin, id, by = Is0B.o, bs = "fs", m = 1)
+ s(timebin, id, by = Is0I.o, bs = "fs", m = 1)
+ s(timebin, pair, by = correct, bs = "fs", m = 1)
+ s(timebin, pair, by = Is1B.o, bs = "fs", m = 1)
+ s(timebin, pair, by = Is1I.o, bs = "fs", m = 1)
+ s(timebin, pair, by = Is0B.o, bs = "fs", m = 1)
+ s(timebin, pair, by = Is0I.o, bs = "fs", m = 1)
```

The p-values for the parametric and smooth difference terms were obtained from the model summary; note that these are only approximations (S. Wood, 2013).

Chapter 3: Results

3.1 Behavioural performance

3.1.1 Overall perception

Figure 4A shows overall behavioural performance, including timed-out responses, across all conditions. Overall, this pattern of results is consistent with the findings of the previous study (Blank et al., 2018): listeners correctly reported most matching written/spoken word pairs as "same" (78.7% ± 4.1%) and mismatching word pairs as "different" (94.6% ± 2.3%). Responses in the partial mismatch condition were more variable: on 37.3% of these trials (SD = 4.9%), participants erroneously perceived that the degraded word they heard matched the expected written word (see also Figure 2, Panels C and D). Some participants were more prone to

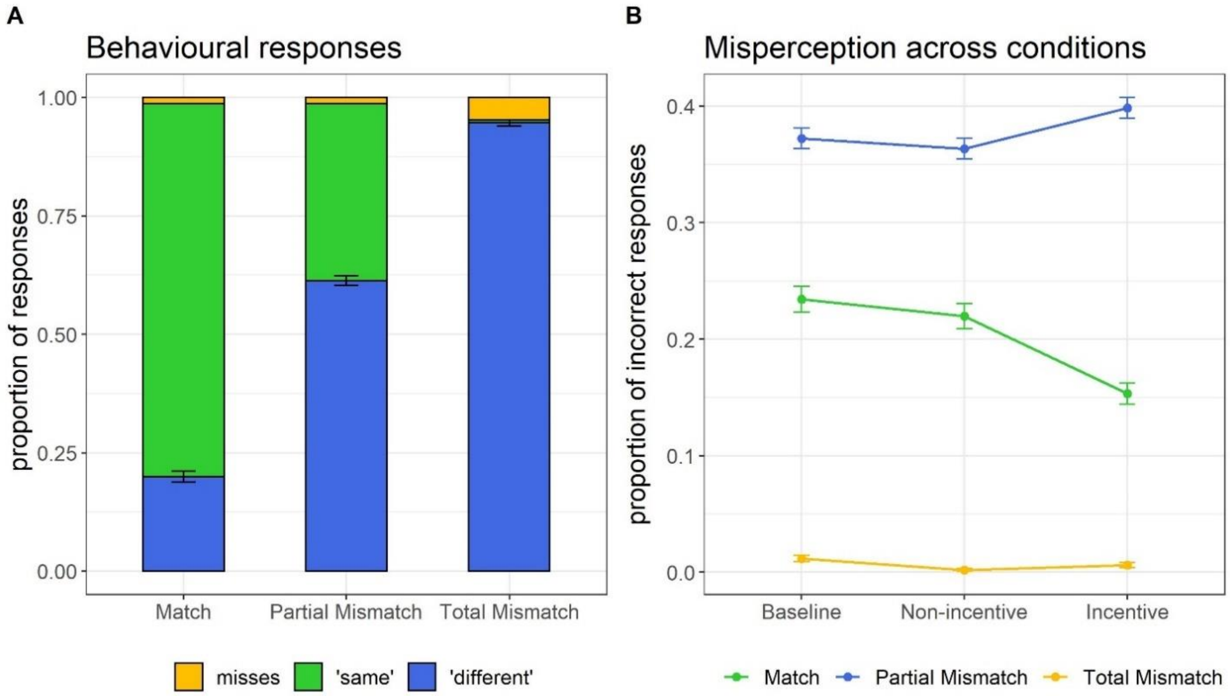


Figure 4. Behavioral performance. (A) Listeners provided more "same" responses in the match than in mismatch or partial mismatch conditions. "Different" responses were more prevalent in the mismatch conditions. Partial mismatch conditions show a large proportion of both "same" and "different" responses, indicating frequent misperception. (B) Perceptual accuracy was lowest in partial mismatch condition (35–40% misperception), followed by match condition (15-23% misperception). Incentive decreased perceptual accuracy on partial mismatch trials but improved perception of matching pairs. Error bars represent SEM.

misperception—others, on the contrary, were deceived less often: individual rates of misperception on partial mismatch trials ranged from as little as 27.8% to as much as 62.9%.

Figure 4B illustrates the interaction between prior knowledge and incentive conditions, focusing specifically on misperception across different types of trials (baseline, incentive, non-incentive). Incentive trials resulted in a lower overall rate of incorrect responses for matching written/spoken pairs (15.3% ± 0.01) and a slightly higher rate of misperception for partial mismatch pairs (39.9% ± 0.01) relative to baseline trials (matching pairs: 23.4% ± 0.01; partial mismatch pairs: 37.2% ± 0.01) and non-incentive trials (matching pairs: 22.0% ± 0.01; partial mismatch pairs: 36.3% ± 0.01) respectively.

To investigate whether incentives had a significant influence on the effects of prior knowledge (i.e., on the likelihood of misperception induced by (mis)matching written text), a mixed logistic model was fitted to all behavioural responses. Wald chi-square tests of effects revealed that there was a significant interaction between prior and reward ($\chi2(4) = 59.7$, $p < 0.000$), as well as a significant main effect of prior knowledge on perception ($\chi2(2) = 112.7$, $p < 0.000$), but no significant main effect of reward ($\chi2(2) = 4.5$, $p = 0.11$). Table 2 reports the results of *post hoc* pairwise comparisons for simple main effects (Bonferroni-adjusted for 12 tests). There were significant differences between matching and mismatching priors, as well as between partially matching and mismatching priors, across all three levels of incentive (all $p < 0.0001$). Both partially mismatching and, interestingly, truly matching written text increased the odds of misperception relative to a clearly mismatching prior that was easy to reject. This suggests that listeners placed relatively low confidence in written text, doubting its reliability even when it was truly informative. Incentives had no clear influence on perceptual outcomes except in one sense: incentive trials appear to have increased listeners' reliance on prior knowledge. This resulted in fewer “misperceptions” triggered by mistrusting a truly matching text ($p < 0.0001$) and a somewhat higher rate of misperception triggered by failing to reject a partially mismatching text. The latter effect did not, however, reach significance in this omnibus model ($p = 0.19$).

*Table 2. Post-hoc pairwise comparisons for the Incentive * Prior Knowledge interaction in a logistic mixed model fitted to all behavioural responses. P-values are Bonferroni corrected.*

Incentive	**Prior**	**Contrast**	**Odds Ratio**	**SE**	**Z ratio**	**P-value**
baseline	.	match / mismatch	42.055	21.795	7.215	<.0001
baseline	.	partial / mismatch	69.116	32.555	8.993	<.0001
non-incentive	.	match / mismatch	223.945	164.904	7.349	<.0001
non-incentive	.	partial / mismatch	365.723	257.364	8.387	<.0001
incentive	.	match / mismatch	43.839	24.942	6.645	<.0001
incentive	.	partial / mismatch	161.953	84.945	9.699	<.0001
.	mismatch	non-incentive / baseline	0.171	0.107	-2.827	0.06
.	mismatch	incentive / baseline	0.509	0.211	-1.627	1
.	match	non-incentive / baseline	0.909	0.089	-0.976	1
.	match	incentive / baseline	0.53	0.056	-6.005	<.0001
.	partial	non-incentive / baseline	0.903	0.066	-1.402	1
.	partial	incentive / baseline	1.192	0.087	2.413	0.19

3.1.2 Perception on partial mismatch trials

Next, we analyzed partial mismatch trials separately, using a logistic mixed model that directly assessed the relative effects of incentives and acoustic similarity on the rate of misperception. Table 3 reports the results of this model fitted using lme4 and brms (full model output can be found in Appendix A for the frequentist model fitted with lme4 and Appendix C for the Bayesian model fitted with brms). The effect of incentives on the perception of partially mismatching pairs was significant ($\chi2 = 13.06$, $p = 0.001$) but its direction was opposite to that predicted. Namely, participants were *more* likely to be deceived into reporting that written and spoken words were “same” on incentive trials than on non-incentive trials ($z = -3.893$, $p = 0.0001$). The odds of misperception were 26% *higher* for incentive trials in the reward block compared to non-incentive trials in the same block (b = -0.30 ± 0.08). On non-incentive trials, the odds of misperception were 15% lower than in the baseline block (b = 0.11 ± 0.08), although these differences in performance were not statistically significant (z = 1.313, p = 0.19). Inter-subject variability for reward was only minor (SD = 0.28 ± 0.12, 95%CrI = [0.05; 0.51]), particularly for incentive vs non-incentive contrast (SD = 0.15 ± 0.10, 95%CrI = [0.01, 0.38]).

Table 3. Results of behavioural analyses with logistic mixed models. Parameter estimates are given in log-odds. B = baseline, I = incentive, N = non-incentive.

	GLM (lme4)	**Bayesian GLM (brms)**
(Intercept)	**0.99** (0.28) ***	**0.99** (0.30) [0.41; 1.60] *
Reward: N vs B	0.11 (0.08)	0.11 (0.09) [-0.06; 0.28]
Reward: I vs N	**-0.30** (0.08) ***	**-0.30** (0.08) [-0.45; -0.14] *
Acoustic similarity (standardized)	**1.26** (0.53) *	**1.18** (0.54) [0.13; 2.26] *

*** p < 0.001, ** p < 0.01, * p < 0.05 (or 0 outside the 95% credible interval for the Bayesian model).

Acoustic similarity between expected and heard words was also predictive of misperception ($\chi2$ = 5.35, p = 0.021). The odds of misperception were 3.5 times higher for written/spoken word pairs whose six-channel vocoded versions were more acoustically similar (b = 1.26 ± 0.53, 95%CrI = [0.13, 2.26]) than for more acoustically distinct words (z = 2.363, p = 0.018). Note that, because of standardization, this coefficient reflects a change in 2 standard deviations, which, in the current dataset, approximates the acoustic difference between the most dissimilar and the “average” word pair—or between the “average” word pair and the most similar one. Indeed, some partial mismatch pairs were consistently misperceived, while others were almost always judged correctly (see also Figure 2C).

Finally, we additionally investigated the effects of potential confounding factors: time-within-experiment (i.e., practice, fatigue, or perceptual learning effects) and reaction time. Neither effect was statistically significant and neither improved model fit (Likelihood ratio test:

$\chi2$ = 0.08, p = 0.78 for trial effects, and $\chi2$ = 1.31, p = 0.25 for reaction time). The estimated effect of time was essentially zero (b = 0.03 ± 0.11), while slower RT had an overall negative effect on accuracy, increasing the odds of misperception by 11% (b = -0.12 ± 0.11, 95%CrI =[0.34, 0.10]). Since neither of these nuisance factors had a significant impact on the pattern of results reported above, these terms were not included in the final model reported in Table 3. The full model output for the nuisance-factor model can be found in Appendix B.

3.1.3 Consistency of responses

We further sought to verify the conclusion of the previous study, namely, that the likelihood of misperception on partial mismatch trials depends on the acoustic similarity of deviating, rather than matching, sounds (Blank et al., 2018). To determine that, we used previously described methods and computed the sum of squared differences between the rate at which each written/spoken pair (e.g., *kit–tit*) was misperceived as "same" with the rate of misperception within its "common sound" group (i.e., three other word pairs that share the same common sounds; here, _it: *tit–kit, wit–lit, lit–wit*) and its "deviating sound" group (i.e., three other word pairs that share the same deviating sounds; here, -k/+t: *kip–tip, Kim–Tim, kin–tin*). This analysis is concerned with the consistency of behavioural responses within each group. Low sum squared difference indicates that the rate of misperception was similar for all items in a group. It means that all four words pairs were either consistently misperceived as "same", or consistently reported as "different", or some mix of the two—as long as responses for all items are similar, the sum squared difference for this group will be low. High sum squared difference means the opposite: that the rate of misperception for one pair tells us little about the perception of other pairs in the group.

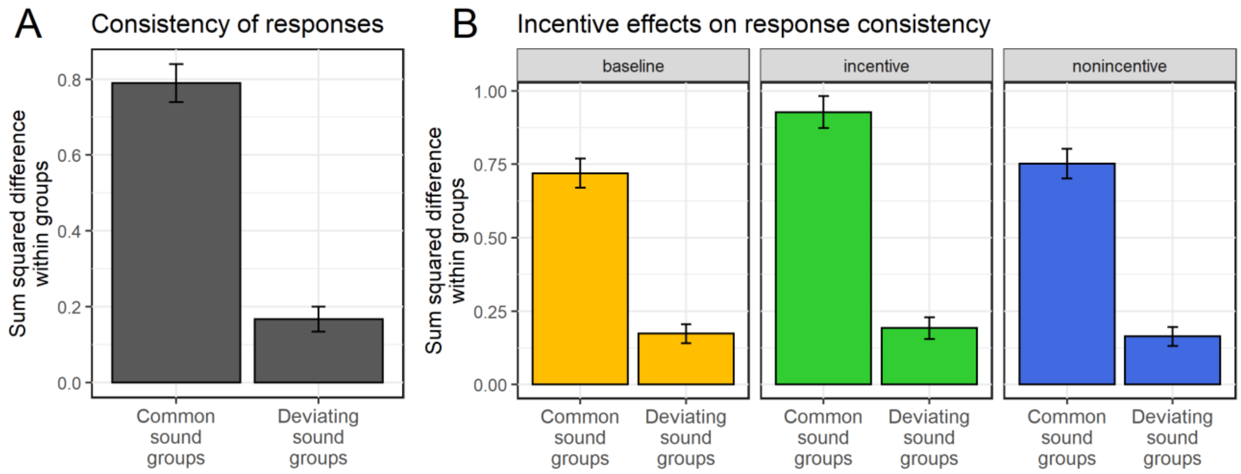


Figure 5. Perceptual outcome on partial mismatch trials is better predicted by the identity of deviating sounds. (A) Mean sum squared differences in the rate of "different" responses for common and deviating sounds groups: responses were more consistent (i.e., lower sum squared difference) for word pairs sharing the same deviating sounds (e.g., -k/+t: kit-tit, kip–tip, Kim–Tim, kin–tin) than for word pairs sharing the same common sounds (e.g., /_it/: kit-tit, tit–kit, wit–lit, lit–wit). (B) Mean sum squared differences for common and deviating sound groups split by incentive condition: incentive did not affect consistency of responses within deviating sound groups but further reduced it within common sound groups. Error bars show the SEM.

As in the previous study, we found that partial mismatch pairs sharing the same deviating sounds had lower sum squared difference than pairs sharing the same common sounds (paired t-test: t(63) = 9.582, p < 0.001); see Figure 5A. This means that perceptual outcomes are better predicted by the sounds that deviate between prior expectations and reality, compared to sounds that are consistent with prior expectations.

Figure 6 provides the rate of misperception within each deviating sound group (Panel A) and within each common sound group (Panel B). While the sum squared differences analysis showed that responses within each deviating sound group were consistent, their perceptual difficulty turned out to be quite different. Some deviating sounds were consistently perceived correctly (see Figure 6A "P/T")—others were more perceptually difficult, leading to more frequent misperceptions (see, e.g., Figure 6A "N/M"). Interestingly, perception of some "mirror" deviating sound groups, such as -k/+p (*cap–pap, cat–pat, can–pan, cam–Pam*) and -p/+k (*pap–cap, pat–cat, pan–can, Pam–cam*), also differed substantially: the former was misperceived

72.6% of the time, the latter—only 13.2% (see Figure 6A “K/P”). This suggests that the perceptual difficulty of a given acoustic contrast might depend on the exact nature of deviating sounds, rather than simply the position (onset vs offset) and type (e.g., N/M vs P/T) of mismatch.

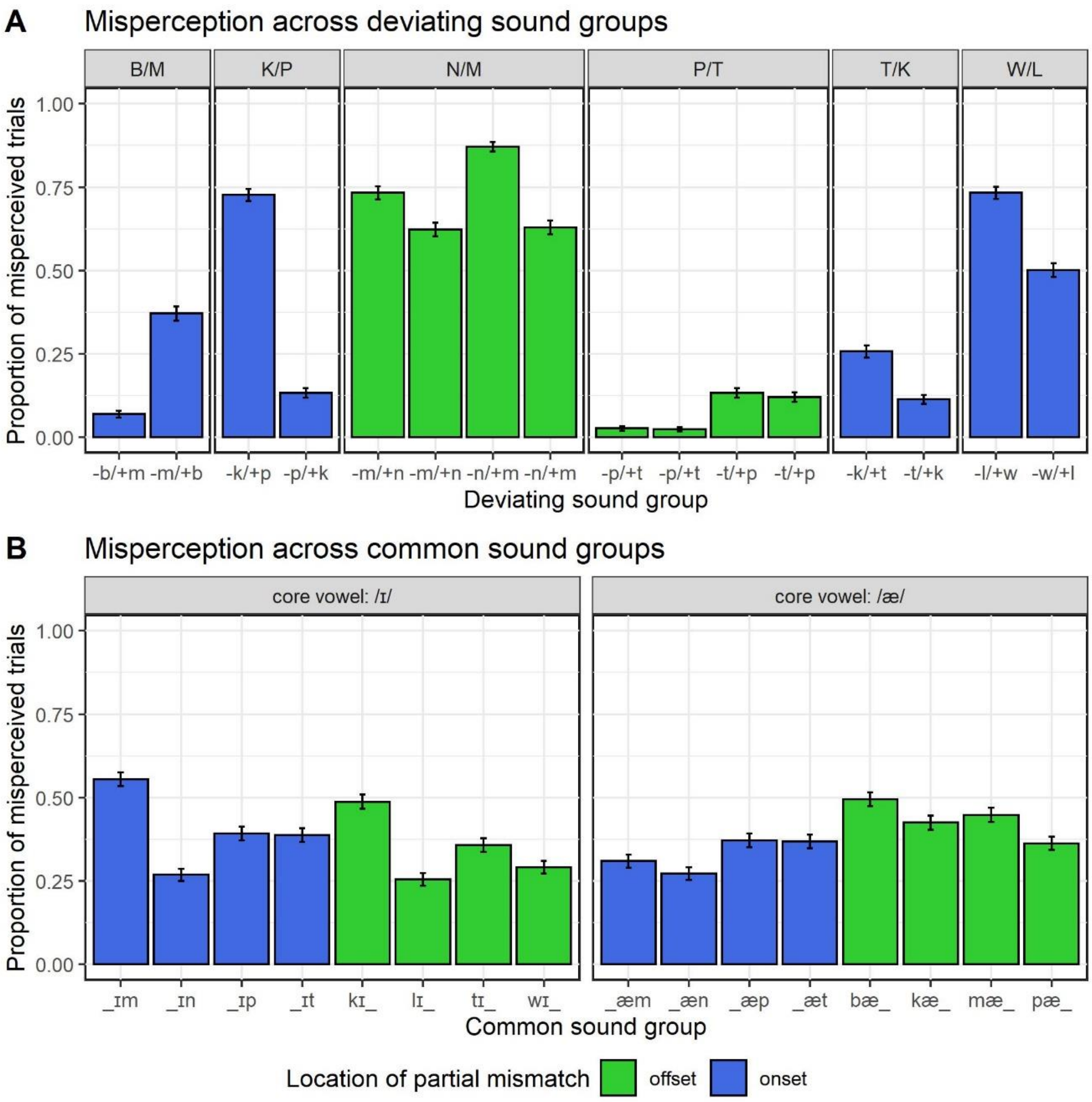


Figure 6. The rate of misperception across all common and deviating sound groups. (A) The rate of misperception varied substantially from one deviating sound group to another, depending both on the type and the exact nature of each acoustic contrast. (B) The mean rate of misperception across common sounds groups was relatively consistent. Error bars show SEM.

To investigate the effect of incentives on the consistency of behavioural performance, we conducted two separate ANOVAs on sum squared differences for common and deviating sound groups. For the common sound group, there was a significant effect of reward: F(2, 189) = 4.718, p = 0.01; for the deviating sound group, this effect was not significant: F(2, 189) = 0.179, p = 0.836. Post-hoc Tukey's test showed that, when a reward was at stake (incentive trials), perception within common sound groups became even less consistent than it was on non-incentive (p = 0.04) and baseline trials (p = 0.01). There was no difference between non-incentive and baseline trials (p = 0.90); see Figure 5B. This is consistent with the overall effect of incentives on the critical perception of the prior in our experiment: on incentive trials, plausible priors (both matching and partially mismatching) were more often accepted as "true" relative to other conditions. More confidence in the prior resulted in a larger proportion of correct responses on truly matching trials (*kit–kit*) but a higher rate of misperception on partial mismatch trials (*kit–tit*).

For the sake of completeness, we ran an additional exploratory analysis investigating whether the effect of motivational incentives depended on the perceptual difficulty of the partial mismatch contrast. Since perceptual outcomes were better predicted by the nature of deviating, rather than common, sounds, we modelled the proportion of misperceived trials within each deviating sound group as a function of the incentive condition (baseline, incentive, non-incentive) and perceptual difficulty (low, high) using a simple linear regression in lme4 (Bates et al., 2015). Perceptual difficulty was estimated from the mean rate of incorrect responses within each deviating sound group. "Low" perceptual difficulty was assigned to groups that were misperceived less than 50% of the time—"high" difficulty was assigned to groups misperceived more often than that. This model showed that while incentives did not affect perception of “easier” sound groups (b = -0.01, SE = 0.03, t = -0.570, p = 0.57), they increased the risk of misperception for more perceptually difficult groups (b = 0.09, SE = 0.04, t = 2.208, p = 0.029); see also Figure 7.

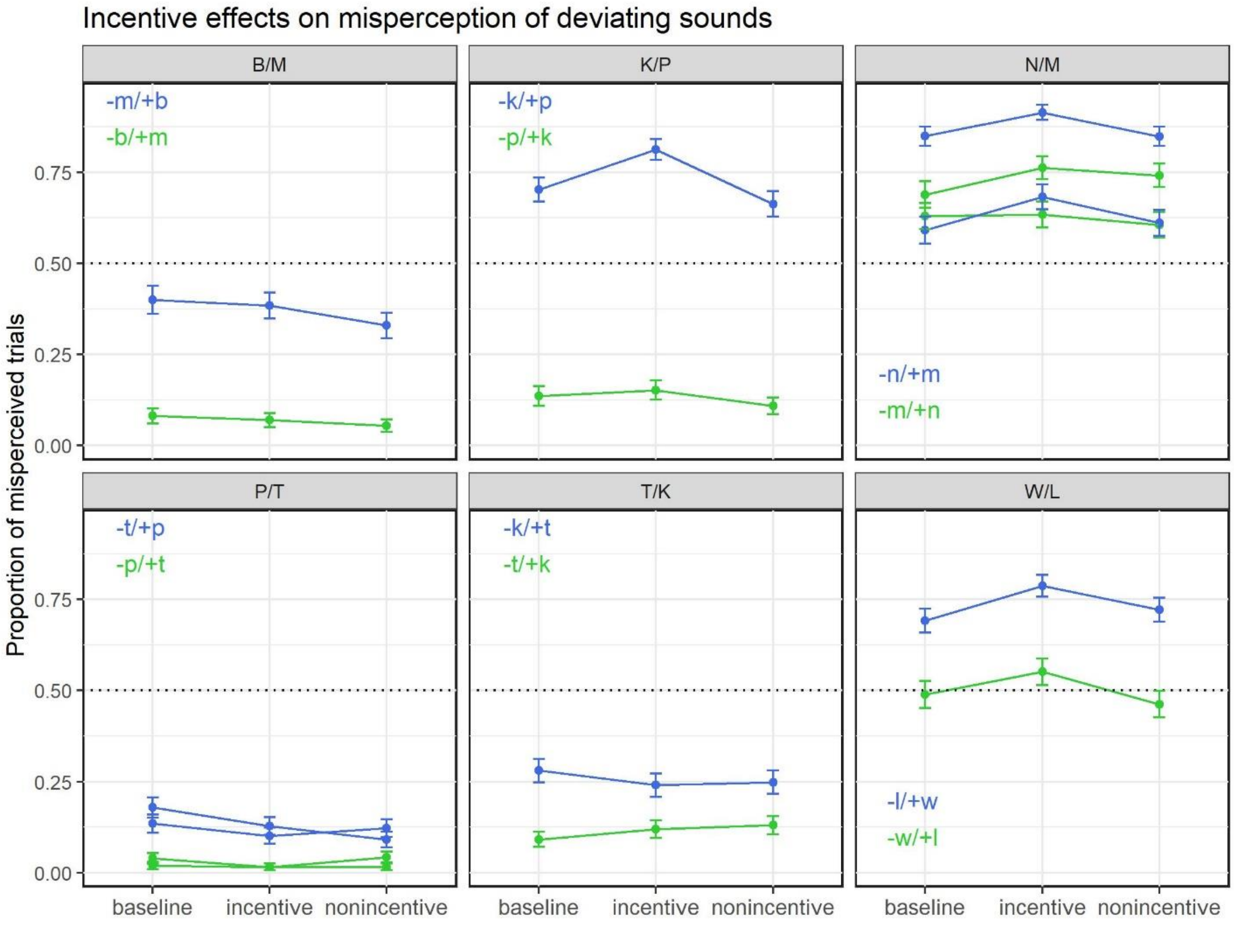


Figure 7. The effects of incentive on misperception of individual deviating sounds groups. Blue represents a more perceptually difficult contrast within each sound group, green—an easier one (cf Figure 6A)

3.2 Pupillometry

Tables 4 and 5 summarize the results of the GAMM model investigating the effects of time and reward on pupil dilation during perception and misperception of partial mismatch pairs (see Appendix D for the full model output). Table 4 gives estimates for the parametric coefficients reflecting the difference in height for correct and incorrect non-incentive trials (non-incentive trials being the reference level of the reward term). Row 2 shows that pupil dilation during misperception and veridical perception of partial mismatch non-incentive pairs was essentially the same. Note, however, that parametric coefficients are not very informative, as they only model the constant difference between conditions (i.e., height adjustment) and account neither for the effects of time within trial nor for the effects of reward. Table 5 reports smooth terms for the same model. Rows 1 and 2 illustrate the significance of non-linear pupil trajectories corresponding to misperceived and correctly perceived non-incentive trials (the two reference curves). Significant difference, in this case, merely means that these two curves are significantly different from 0. Our main interest, however, is the difference between non-incentive and baseline trials (corresponding to the block effect of reward), and that between non-incentive and incentive trials (the trial effect of reward), so we focus on Rows 3 to 6 instead. These correspond to smooth parameters that estimate how pupil trajectories change over time as a result of different reward conditions during perception and misperception of mismatch. Specifically, Rows 3 and 5 indicate that the difference over time between baseline and non-incentive trials (i.e., the block effect of reward) was significant for both accurately perceived (F = 1.99, p = 0.04) and misperceived word pairs (F = 3.43, p = 0.03). The estimated degrees of freedom, or edfs, represent an estimate of the wiggliness of the pattern. Higher edf estimates correspond to more wiggly (more complex) smooths. Here, we see that this difference is rather complex during veridical perception but is close to linear during misperception. In a similar vein, Rows 4 and 6 show that the difference between pupil dilation trajectories on non-incentive vs incentive trials (i.e., the trial effect of reward) was significant for both perceptual outcomes: F = 2.85, p = 0.002 for correctly perceived trials and F = 2.18, p = 0.04 for misperceived trials. As before, the low edf number indicates that the non-linear pattern during misperception is less complex relative to veridical perception, although this difference is not as dramatic as before. Still, this lack of complexity in the shape of the “misperception” curves suggests that the difference between conditions during misperception is constant, rather than non-linear.

Table 4. Parametric coefficients of the generalized additive mixed model on perceived and misperceived partial mismatch trials, across the three levels of incentive (baseline, non-incentive, incentive).

Parametric terms	Estimate	SE	t-value	p-value	
Intercept (misperceived N)	0.225	0.059	3.827	0.0001	***
correct N	-0.072	0.081	-0.893	0.372	

*Table 5***.** *Smooth function terms of the generalized additive mixed model on perceived and misperceived partial mismatch trials, across the three levels of incentive (B = baseline, N = non-incentive, I = incentive). The first two lines show the smooth terms for the reference level (non-incentive trials). Lines 3–6 represent binary smooths—smooth difference curves comparing pupil trajectories for incentive and baseline trials against the non-incentive reference curves—both for correctly perceived (lines 3–4) and misperceived (lines 5–6) trials. Note that binary smooth factors collapse constant and smooth difference into a single term. The edf column shows the estimated degrees of freedom, reflecting the wiggliness, or complexity, of each curve. The maximum allowed wiggliness (controlled by the k-parameter in model settings) was set to 19 for the reference curves and to 9 for the difference curves.*

Smooth terms	edf	F-value	p-value	
s(timebin) : misperceived (N)	18.008	38.241	< 0.000	***
s(timebin) : correct (N)	18.177	56.35	< 0.000	***
s(timebin) : correct B vs correct N	8.729	1.988	0.040	*
s(timebin) : correct I vs correct N	7.763	2.851	0.002	**
s(timebin) : misperceived B vs misperceived N	2.007	3.416	0.032	*
s(timebin) : misperceived I vs misperceived N	4.463	2.184	0.039	*

To investigate whether this is indeed the case, we followed up this analysis with an analogous GAMM fitted using ordered factors. The results of this model are summarized in Tables 6 and 7. Parametric coefficients in Rows 3 and 5 of Table 6 indicate that there is a significant difference in height between baseline and non-incentive curves, both during perception ($t = 2.13$, $p = 0.03$) and misperception ($t = 2.24$, $p = 0.02$). Coefficients in Rows 4 and 6 of Table 6 indicate that the same holds true for the incentive vs non-incentive difference, during perception ($t = 2.68$, $p = 0.007$) and misperception ($t = 1.98$, $p = 0.048$) alike. This pattern of results suggests that pupil diameter during baseline and incentive trials is larger than during non-incentive trials, regardless of the perceptual outcome. Turning to smooth terms, reported in Table 7, the shape of pupil trajectory over correct incentive trials is significantly different from the dilation pattern observed over correct non-incentive trials (F = 2.56, p = 0.008). A relatively high number of edfs suggest that this difference wave is also rather complex in shape. At the same time, there are no significant non-linear differences between pupil trajectories for baseline vs non-incentive trials (both perceived and misperceived) or during misperceived incentive vs non-incentive trials (all p > 0.05).

Table 6. Parametric coefficients of the ordered factor GAM model investigating whether the difference between the pupil trajectories during (mis)perception was constant or non-linear.

Parametric terms	Estimate	SE	t-value	p-value	
Intercept (misperceived N)	0.224	0.059	3.809	0.0001	***
correct N	-0.072	0.081	-0.88	0.379	
correct B vs correct N	0.131	0.061	2.132	0.033	*
correct I vs correct N	0.168	0.062	2.684	0.007	**
misperceived B vs misperceived N	0.175	0.078	2.244	0.025	*
misperceived I vs misperceived N	0.164	0.083	1.976	0.048	*

Table 7. Smooth function terms of the ordered factor GAM model investigating whether the difference between the pupil trajectories during (mis)perception was constant or non-linear. As in the previous model, lines 1–2 describe to the reference smooths (pupil dilation pattern during non-incentive trials), while lines 3-6 show the difference wave between incentive–non-incentive (lines 4 and 6) and baseline–non-incentive (lines 3 and 5) pupil trajectories.

Smooth terms	edf	F-value	p-value	
s(timebin) : misperceived (N)	18.009	36.662	< 0.000	***
s(timebin) : correct (N)	18.177	56.806	< 0.000	***
s(timebin) : correct B vs correct N	7.736	1.637	0.098	
s(timebin) : correct I vs correct N	6.748	2.561	0.008	**
s(timebin) : misperceived B vs misperceived N	5.260	0.495	0.81	
s(timebin) : misperceived I vs misperceived N	3.867	1.132	0.29	

Visualization allows for a more intuitive interpretation of these parameters. Figure 8 illustrates these non-linear differences for correctly perceived (Panels B1 and C1) and misperceived (Panel B2 and C2) trials. Areas of significant differences between pupil trajectories over time are highlighted in red. Note that significant difference over a small period of time does not imply that the *entire* corresponding difference wave is significant. Model summary (Table 7) indicates that there is only one significant non-linear difference—between correct incentive and non-incentive trajectories (Panel C1). All other panels simply illustrate *where in time* pupil dilation patterns start to significantly diverge between conditions. Panel A shows fitted effects for this model, or the estimated pupil trajectories for each condition. Note that these are the summed effects, as they include the intercept. Random effects are set to zero.

Back to difference curves, Panels B1 and B2 of Figure 8 show the difference between pupil trajectories during baseline vs non-incentive trials (the block effect of reward), separately for each perceptual outcome. As can be seen on Panel B1, pupil is more dilated during *correct* baseline trials over the pre-stimulus interval (1697-1939 ms into the trial, just before the onset of the spoken word). Panel B2 shows that this pattern changes only slightly during misperception: pupil again is more dilated during the pre-stimulus interval (1373-2262 ms) of baseline vs non-incentive trials. (Note that while ostensibly we pit baseline trials against non-incentive trials, we are essentially interested in the opposite pattern—how non-incentive trials differ from baseline trials. The binary difference smooths on Figure 8 Panels B1 and B2, then, should be interpreted in reverse: pupil is more *constricted* over 1697-1939 ms interval (for correctly perceived pairs, Panel B1) and over 1373-2262 ms interval (for misperceived pairs, Panel B2) during non-incentive trials relative to baseline trials. Modeling the reverse pattern was necessary for estimating trial- and block- effects of reward with a single reference curve.) Figure 8 Panels C1 and C2 illustrate respective differences for incentive vs non-incentive trials. We can see that dilation amplitude rises for incentive relative to non-incentive trials over a period right after the

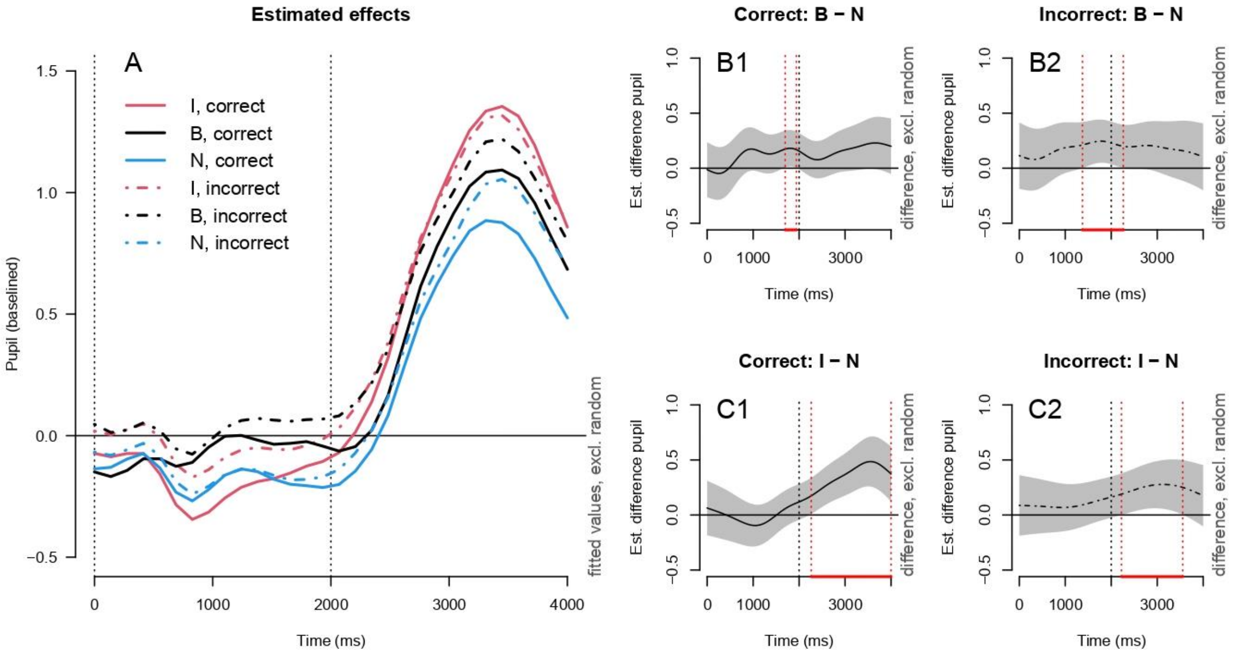


Figure 8. Estimates of the ordered factor GAMM fit to pupil data. Left: Summed effects for all conditions (random effects set to zero). Right: Ordered factor difference curves derived from the model, with pointwise 95% confidence intervals. Significant differences (deviations from the 0 line) are marked in red. 0 ms mark correspond to the onset of written text, 2000 ms—t0 the onset of degraded word. I = incentive; B = baseline; N = non-incentive.

presentation of spoken word (2262-4000 ms for correct trials, 2222-3555 ms for incorrect trials). Because of the latency of the pupil dilation response, the earliest pupillary effects develop at least 220 ms after the manipulation that induced them (Mathôt et al., 2018). Thus, the main trial effect of incentive appears to lie in the reactive, or transient, engagement of control, as opposed to the proactive, or preparatory one. The latter would be expected to occur during the pre-stimulus interval—before the presentation of degraded spoken words. Interestingly, misperception is associated not only with a shortened time window of differences but also with lower dilation magnitude. All in all, phasic differences between incentive and non-incentive dilation trajectories appear to be less dramatic during misperception.

To test whether this was the case and better understand how incentives affected pupil dilation patterns during perception and misperception, we fitted another ordered factor GAM model. This model compared pupil trajectories for correctly perceived vs misperceived trials separately for each level of incentive. The results of this model are reported in Table 8 (parametric terms) and Table 9 (smooth terms). Here, we focus specifically on Rows 4–6 of each table, as these describe the differences between pupil dilation trajectories corresponding to accurate and inaccurate perception for each type of trial. None of these terms reached statistical significance. Figure 9 illustrates non-linear patterns that model the difference between pupil trajectories during perception and misperception at each level of incentive. While there are interesting differences in the overall patterns across reward conditions, the differences curves for the Correct–Misperceived contrasts do not in themselves significantly differ from zero. In other words, while pupil trajectories associated with each perceptual outcome vary across baseline, incentive, and non-incentive conditions, as revealed by the previous model, there is a common pupil dilation pattern that models the difference between correctly perceived and misperceived trials. This difference wave does not appear to be influenced by incentive manipulations.

Table 8. Parametric coefficients of the ordered factor GAM model investigating the pupil trajectories for correctly perceived and misperceived trials differ within each incentive condition. Line 1 refers to the height of the reference curve (pupil dilation pattern during misperceived non-incentive trials), lines 2–3 model the constant difference between baseline–non-incentive and incentive–non-incentive pupil trajectories during misperception. Lines 4–6 represent the difference in height for pupil dilation patterns during veridical perception and misperception—separately for non-incentive, baseline and incentive trials.

Parametric terms	Estimate	SE	t-value	p-value	
Intercept (misperceived N)	0.219	0.053	4.119	<0.000	***
misperceived B vs misperceived N	0.215	0.071	3.022	0.003	**
misperceived I vs misperceived N	0.185	0.071	2.604	0.009	**
correct N vs misperceived N	-0.057	0.074	-0.77	0.441	
correct B vs misperceived B	-0.140	0.080	-1.764	0.078	.
correct I vs misperceived I	-0.062	0.068	-0.919	0.358	

Table 9. Smooth function terms of the ordered factor GAM model investigating the pupil trajectories for correctly perceived and misperceived trials differ within each incentive condition. Line 1 describe to the reference curve (pupil dilation pattern during misperceived non-incentive trials), lines 2–3 correspond to the non-linear difference waves between baseline–non-incentive and incentive–non-incentive pupil trajectories during misperception. Lines 4–6 represent the non-linear difference between pupil dilation trajectories during veridical perception and misperception—separately for non-incentive, baseline and incentive trials. None of these smooths were significantly different from 0 or exhibited substantial wiggliness (as evident from low edf values)

Smooth terms	edf	F-value	p-value	
s(timebin) : misperceived (N)	18.311	77.494	<0.000	***
s(timebin) : misperceived B vs misperceived N	7.634	1.654	0.084	.
s(timebin) : misperceived I vs misperceived N	5.843	2.24	0.032	*
s(timebin) : correct N vs misperceived N	2.986	1.095	0.350	
s(timebin) : correct B vs misperceived B	1.55	0.597	0.536	
s(timebin) : correct I vs misperceived I	3.023	0.326	0.875	

The deviance explained by this and previous models is 12.3%. Although this estimate is relatively low, this is due to the large inter-individual variability common to pupillometric data.

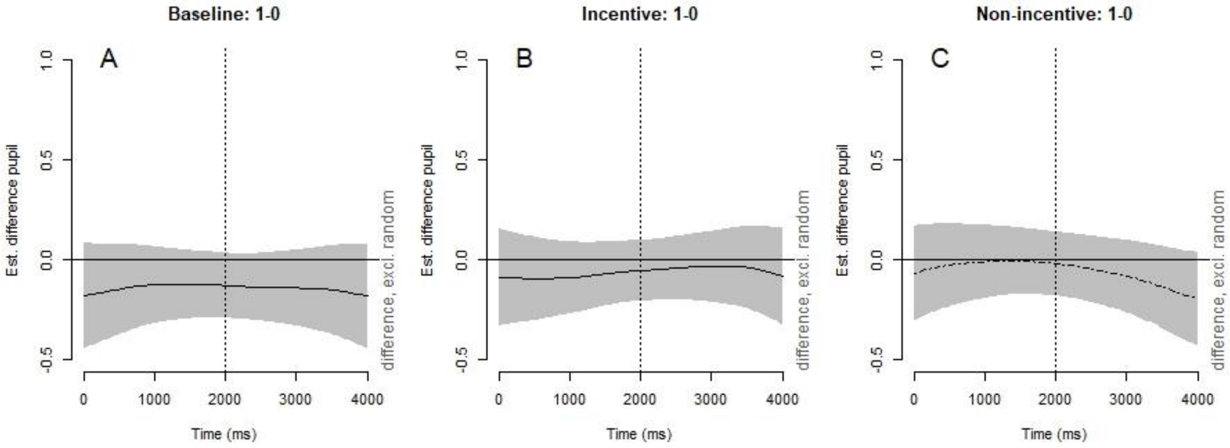


Figure 9. Ordered factor difference curves derived from the GAMM modeling correct–incorrect pupil trajectories within each level of incentive. Gray shading reflects pointwise 95% confidence intervals. All three CIs contain 0 across the entire time series, suggesting that pupil dilation trajectories associated with each perceptual outcome do not significantly differ.

Chapter 4: Discussion

Listening conditions of our daily life are rarely perfect, and mishearing is a common occurrence in everyday conversation. While most slips of the ear are relatively benign and can be easily resolved on the spot—often with a shared amusement,—other perceptual blunders may lead to serious misunderstanding. Hearing-impaired individuals are particularly prone to frequent misperception. Recurring communication breakdowns and chronically elevated listening effort associated with frequent misperception put their social lives at substantial risk, with predictably negative consequences for general health and cognitive well-being (Pichora-Fuller et al., 2015). In this study, we investigated several factors that affect misperception of degraded speech, including (mis)informative prior expectations, the degree of acoustic similarity between expected and heard words, and the level of attentional engagement during listening. Previous evidence suggests that increased attention and cognitive control are both crucial and beneficial for comprehension of degraded speech. However, these effects were never tested in interaction with prior knowledge. At the same time, while it is well-established that misinformative expectations are conducive to misperception—especially when they are acoustically close to the degraded target,—it is less clear to what extent such expectation-induced misperceptions may be due to inattentive listening.

Here, we were able to test this hypothesis in a rewarded version of the same/different task, using pupillometry as a proxy of moment-to-moment attentional engagement. We used written text to create prior expectations about the upcoming degraded spoken word, thus "priming" listeners to misperceive it for the visually presented word. We additionally varied the relative importance of correct perception on each trial via motivational incentives in order to manipulate listeners' attentional engagement on a trial-to-trial basis. In addition to behavioural report, we tracked listeners' pupil size to assess the effect of incentives and prior knowledge on perceptual processing that leads up to each perceptual outcome (correct perception vs misperception).

4.1 (Mis)informative prior expectations trigger misperception

Our behavioural results were generally in line with previous studies (Blank et al., 2018; Sohoglu et al., 2014). Listeners confidently rejected misinformative priors on total mismatch trials (*kit–ban*), slightly less confidently accepted informative priors on total match trials (*kit–kit*), and were often deceived to report "same" on partial mismatch trials (*kit–tit*). We also confirmed the previous observation that the perceptual outcome for a given partial mismatch pair (such as *kit–tit*) is better predicted by the perception of word pairs that share the same deviating sounds (in this example, -k/+t onset: *kip–tip*, *Kim–Tim*, *kin–tin*) as opposed to word pairs that share the same common sounds (i.e., /-it/ offset: *tit–kit*, *wit–lit*, *lit–wit*).

At the same time, while the overall rate of misperception on partial mismatch trials (37.3%) was close to that reported by Blank and colleagues (~40%), we observed a much higher percentage of misperception on total match trials (20%,—vs less than 10% reported in the previous study). This two-fold increase in the likelihood of misperception persisted across all three levels of incentive (ranging from 15% on incentive trials to 23% on baseline trials). We attribute this discrepancy to the differences between our datasets and experimental setups. Our task required a categorical same-or-different response—Blank and colleagues (2018), on the other hand, asked for a more nuanced report that additionally included confidence judgement (e.g., "definitely same" vs "possibly same", "definitely different" vs "possibly different"). It is possible that this task setup allowed them to better accommodate response uncertainty, leading to a more accurate report on matching trials. More importantly, though, these findings indicate that even *informative* prior knowledge is not always beneficial to perception. Whether prior expectations are employed to inform perceptual inference or discarded in favor of sensory signal appears to depend on their expected validity and the task set.

In both studies, global, experiment-wise validity of the prior was relatively low—the written text matched spoken words on just one-quarter of trials, with half of the written/spoken pairs varying in only one segment. The datasets were also small (each comprised of just 32 words), and participants heard—and saw—each word multiple times throughout the experiment, becoming increasingly aware that the task set was fixed. Thus, both groups of listeners have, in all likelihood, strategically lowered their confidence in written priors to avoid being tricked. In addition, the same/different task itself essentially pitted prior expectations against sensory input, asking listeners to decide between trusting their ears (that heard barely intelligible speech) and going with their predictions (which were invalid 3 out of 4 times). In current experimental settings, this was akin to choosing the lesser of two evils. Such a perceptual dilemma is unlikely to occur in real listening situations that require open-set word recognition rather than simple discrimination. In open-set identification, prior expectations—even if mildly misinformative—help to acoustically *approximate* the target, while perhaps hindering listeners' ability to *home in* on the exact version of what was said. Word recognition "task" is rarely performed in isolation: some sentence- or situational context is always available to support speech perception. Furthermore, the sheer diversity and unpredictability of real-life listening scenarios make it harder to compute the expected "validity" statistics for prior expectations and adjust one's confidence in priors. Thus, the discounting of prior knowledge observed here is likely an artifact of the experimental setup. In real listening situations, prior expectations (e.g., from the semantic context) are not only relied upon but often applied inflexibly. Older adults and hearing-impaired individuals, in particular, show an increased bias to respond consistently with the context (Rogers et al., 2012; Rogers & Wingfield, 2015; Signoret & Rudner, 2019). They also report higher confidence when responding in line with contextual cues, even when it results in misperception (Rogers et al., 2012; Rogers & Wingfield, 2015).

4.2 Incentives increase reliance on prior knowledge

Incentives did influence perception, but not in the expected direction. Trial-level incentives not only failed to improve perceptual inference on partial mismatch trials but made it worse. When offered a bonus for accurate perception, listeners were 26% more likely to misreport that written and spoken words were "same" relative to when they had no knowledge of reward (baseline trials) or had no particular incentive to exert additional effort during listening (non-incentive trials). Concurrently, we observed a significant increase of "same" responses for total match pairs—which, in this case, actually improved perceptual accuracy. This pattern of results is uniquely consistent with an account of speech perception where improved attentional engagement increases reliance on prior expectations. From the predictive coding perspective then, incentives appear to have increased attention to the written text, essentially improving precision of predictions rather than enhancing sensitivity to the upcoming auditory input. Thus, if degraded speech signal is not sufficiently informative to generate a strong prediction error signal, the prior is accepted. Notably, this strategy makes a lot of sense for the perceptual discrimination task employed here. In the same/different judgement task, the prior remains task-relevant regardless of its informativeness (i.e., even if mismatching). Unfulfilled predictions straightforwardly bias listeners toward a correct decision: the degraded spoken word is "different", even if it cannot be accurately recognized, as long as the signal contains sufficient acoustic information to cast a doubt on the (mis)informative prior. Indeed, as demonstrated in

Figure 7, incentives increased the risk of perceptual confusion for more difficult word pairs, without affecting perception of easier discriminable words. The outcomes of total mismatch trials were also not affected by motivational manipulation—when the written text clearly mismatches the auditory signal, listeners reliably make accurate perceptual decisions.

Thus, while incentives increase reliance on prior knowledge in more ambiguous cases, they do not affect perceptual inference when the sensory signal is sufficiently strong to override misinformative prediction. This response is consistent with the inverted U-shaped relationship between cognitive control and task difficulty, predicting greater allocation of cognitive resources on more challenging listening conditions (Eckert et al., 2016; Poldrack et al., 2001; Zekveld et al., 2006). Although all words in our study were noise-vocoded with the same number of bands, prior knowledge and acoustic similarity between written and spoken words directly affected the perceptual difficulty of the task. As evident from the behavioural report, total mismatch trials were the easiest, while partial mismatch trials with acoustically similar word pairs were the most challenging. And it is at this highest difficulty level that the effect of incentives became apparent. In fact, similar results were reported by Richter (2017) for a rewarded tone discrimination task, which was conceptually analogous to ours. In that study, participants listened to sequences of tones that were either identical, differing by 3Hz or differing by 20Hz. They then reported whether these tones were the same or different. Monetary reward increased exerted listening effort, but only in a difficult condition—when participants performed discrimination in a block that consisted of identical and 3Hz trials. On the other hand, several studies that failed to detect the behavioral effects of incentives during listening in noise also failed to investigate these effects across a sufficient range of difficulty levels (Koelewijn et al., 2018, 2021). Admittedly, our study was also limited in this respect, since the small number of stimuli made it challenging to probe the interaction between incentives and acoustic similarity (listening demand) beyond (mis)perception of individual word pairs. Overall, however, our results are entirely consistent with the neuroeconomic account of cognitive control (Brehm & Self, 1989; Shenhav et al., 2017), in which additional cognitive resources are allocated to accommodate increasing task demands, as long as the costs of engaging control do not outweigh its benefits.

Effortful speech comprehension in challenging listening conditions frequently evokes elevated activity in cingulo-opercular and frontoparietal attentional networks (Adank et al., 2012; Alain et al., 2018; Erb & Obleser, 2013; Hervais-Adelman et al., 2012; Ritz et al., 2021; Vaden et al., 2013). This neural activity is often interpreted as reflecting a top-down compensatory mechanism. It is thought to engage a range of predictive processes that use lexical, syntactic, and contextual cues to compensate for the impoverished auditory encoding and correspondingly imprecise representations of speech in auditory short-term memory. In the context of degraded sentences, these predictive processes often benefit speech comprehension, particularly when the sentence context itself is sufficiently rich (Obleser et al., 2007; Obleser & Kotz, 2010; Rysop et al., 2021). In the context of isolated words, however, top-down contextual predictions are both relatively useless and difficult to generate. Instead, listeners appear to rely on abstract phonological representations of previously presented speech stimuli (priors), using these representations as high-precision templates against which to compare degraded sensory inputs. Motivational incentives enhance cognitive control, further increasing reliance on these top-down predictive mechanisms. In our task, this strategy turned out to be of questionable utility. While higher confidence in prior knowledge clearly improved perceptual accuracy on total match trials, it promptly backfired by worsening perception of partial mismatch trials. It is possible that adjusting the intensity of cognitive control via incentives could be more efficacious in tasks that directly benefit from increased use of contextual, semantic, and other higher-order linguistic cues—such as sentence comprehension or listening to a coherent story.

4.3 Acoustic similarity measure fails to capture perceived acoustic differences

The analysis of response consistency within common and deviating sound groups revealed that, while incentives did not affect perceptual processing of word pairs sharing the same deviating sounds (e.g., *kit–tit, kip–tip, Kim–Tim, kin–tin*), they decreased the consistency of responses within common sound groups (e.g., *kit–tit, tit–kit, wit–lit, lit–wit*). This pattern of results is likely driven by "mirror" written/spoken pairs (such as *kit–tit* and *tit-kit*). Perception of such pairs differed quite substantially, as listeners easily rejected the prior in one direction (e.g., -k/+p) but could hardly detect the opposite difference (-p/+k); see Figure 6A. Unfortunately, our acoustic similarity measure was not sensitive to such differences: gammatone spectral analysis simply compared spectral profiles of noise-vocoded words that comprised each written/spoken pair. Thus, pairs *cat/pat* and *pat/cat*, for example, received the same acoustic similarity ratings—but listeners perceived them very differently. The former was always misperceived as "same" (misperception rate: 97.1%), the latter—as "different" (misperception rate: 0.8%); see also Figure 2C. Despite this fact, acoustic similarity still had a large impact on perceptual outcomes driving a 3.5-fold increase in the likelihood of misperception between the most acoustically dissimilar partial mismatch pair and an "average" pair. In comparison, motivational incentives accounted for just 1/13th of this effect size. Nonetheless, future research might benefit from finding a more precise metric of acoustic similarity that directly reflects *perceived* acoustic differences between degraded tokens.

4.4 Incentives affect perception via reactive and proactive control

Our GAMM analyses showed that motivational incentives strongly affected pupil dilation trajectories during perceptual decision-making on partial mismatch trials. First, we observed a constant difference in height of pupil trajectories associated with each type of trial (baseline, incentive, non-incentive). Specifically, when listeners were intrinsically motivated (baseline trials) and extrinsically motivated (incentive trials), their pupil was consistently more dilated relative to periods of mild disengagement (non-incentive trials). Non-incentive trials, in general, were associated with low tonic/low phasic dilation pattern. Second, incentive trials additionally showed a dramatic increase in phasic dilation in comparison to non-incentive trials. Importantly, these transient incentive effects were present only after 2200 ms, i.e., in response to the presentation of degraded words. This high tonic/high phasic pattern of results suggests that incentive-related control mechanisms were operating in both proactive and reactive fashion. Proactive, or preparatory, control is implemented in anticipation of the target and requires maintenance of task-related information in working memory. Reactive mechanisms, on the other hand, engage in response to changing task demands and flexibly adjust control on a case-by-case basis. Interestingly enough, there was no difference in phasic dilation between misperceived and correctly-perceived incentive trials, suggesting that implementing control "just in time" does not improve perceptual inference. In fact, veridical perception was generally associated with lower *tonic* dilation. Notably, this pattern persisted across all levels of incentives (see Figure 9)—highlighting the fact that enhanced cognitive control had rather counter-productive effects on our perceptual judgment task.

These findings are not entirely consistent with the predictions of the adaptive gain theory of LC-NE (locus coeruleus-norepinephrine) function (Aston-Jones & Cohen, 2005). According to AGT, the LC-NE system, considered to be the main driver of cognitive pupillary response (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Murphy et al., 2011, 2014), operates via two modes of function: phasic and tonic. The adaptive gain theory posits that elevated tonic activity without pronounced phasic bursts (reflected in larger baseline pupil size) corresponds to distractible attentional state and task disengagement. Phasic activity (reflected in transient, stimulus-driven pupil dilation), on the contrary, facilitates goal-driven behaviours and enhances within-task performance. Here, we observed the opposite pattern: intrinsically- and extrinsically-driven attentional engagement were both associated with greater tonic dilation, while the lack of monetary incentives conversely drove tonic activity down. And while monetary incentives did lead to a profound increase in transient activity, this hardly improved within-task performance. At the same the time, these results are in line with Chiew and Braver (2013, 2014), who found a similar high tonic/high phasic response profile during incentive trials. It is possible that these effects are driven by other neurotransmitter systems known to affect pupil size (Costa, 2016; de Gee et al., 2014; Naicker et al., 2016; Reimer et al., 2016). Since our task involved reward processing, it could have engaged dopaminergic activity, thus altering the expected noradrenergic response dynamics predicted by the AGT and raising these rather complex interactions. Future work might benefit from explicitly modeling these interactions.

4.5 Summary

In this study, we investigated how increased attentional engagement might affect the likelihood of prediction-induced misperceptions. Listeners performed a perceptual decision task with degraded speech stimuli under three levels of incentive: baseline, incentive, non-incentive. We induced frequent perceptual confusion by presenting listeners with noise-vocoded words preceded by matching, mismatching and partially mismatching text. Contrary to our predictions, listeners increased their reliance on prior expectations when a reward was at stake. This strategy resulted in a higher rate of misperception when prior expectations were plausible yet incorrect but improved perceptual accuracy when predictions were truly informative. These incentive-related control processes operated in both the reactive and the proactive modes of function, engaging both sustained attentional control and “just in time” attention. In sum, our work indicates that increased attention is not always beneficial during effortful listening. Recruitment of cognitive control processes increases reliance on prior knowledge when sensory detail is insufficient, which only exacerbates the problem of prediction-induced mishearing—at least in auditory discrimination tasks with isolated words. Future work should focus on investigating these effects on listening tasks with more linguistically complex stimuli, as these are expected to directly benefit from the use of higher-order predictive processes observed here.

References

Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, *50*(1), 77–84. https://doi.org/10.1016/j.neuropsychologia.2011.10.024

Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *39*(7), 2695–2709. https://doi.org/10.1002/hbm.24031

Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., & Millman, R. E. (2020). Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy. *International Journal of Audiology*, *0*(0), 1–11. https://doi.org/10.1080/14992027.2020.1853262

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Auksztulewicz, R., & Friston, K. (2015). Attentional Enhancement of Auditory Mismatch Responses: A DCM/MEG Study. *Cerebral Cortex*, *25*(11), 4273–4283. https://doi.org/10.1093/cercor/bhu323

Bakkour, A., Morris, J. C., & Dickerson, B. C. (2009). The cortical signature of prodromal AD: Regional thinning predicts mild AD dementia. *Neurology*, *72*(12), 1048–1055. https://doi.org/10.1212/01.wnl.0000340981.97664.2f

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beck Lidén, C., Krüger, O., Schwarz, L., Erb, M., Kardatzki, B., Scheffler, K., & Ethofer, T. (2016). Neurobiology of knowledge and misperception of lyrics. *NeuroImage*, *134*, 12–21. https://doi.org/10.1016/j.neuroimage.2016.03.080

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., & Carlyon, R. P. (2013). Lexical Influences on Auditory Streaming. *Current Biology*, *23*(16), 1585–1589. https://doi.org/10.1016/j.cub.2013.06.042

Blank, H., Biele, G., Heekeren, H. R., & Philiastides, M. G. (2013). Temporal Characteristics of the Influence of Punishment on Perceptual Decision Making in the Human Brain. *Journal of Neuroscience*, *33*(9), 3939–3952. https://doi.org/10.1523/JNEUROSCI.4151-12.2013

Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Blank, H., Spangenberg, M., & Davis, M. H. (2018). Neural Prediction Errors Distinguish Perception and Misperception of Speech. *Journal of Neuroscience*, *38*(27), 6076–6089. https://doi.org/10.1523/JNEUROSCI.3258-17.2018

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.55). http://www.praat.org/

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bond, Z. S. (2005). Slips of the Ear. In *The Handbook of Speech Perception* (pp. 290–310). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470757024.ch12

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology*, *66*, 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Boudewyn, M. A., & Carter, C. S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148–155. https://doi.org/10.1016/j.neuropsychologia.2018.05.024

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*, 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Carolan, P. J., Heinrich, A., Munro, K. J., & Millman, R. E. (2021). Financial reward has differential effects on behavioural and self-report measures of listening effort. *International Journal of Audiology*, *60*(11), 900–910. https://doi.org/10.1080/14992027.2021.1884907

Chiew, K. S., & Braver, T. (2013). Temporal Dynamics of Motivation-Cognitive Control Interactions Revealed by High-Resolution Pupillometry. *Frontiers in Psychology*, *4*, 15. https://doi.org/10.3389/fpsyg.2013.00015

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 509–529. https://doi.org/10.3758/s13415-014-0280-0

Chiew, K. S., & Braver, T. S. (2016). Reward favors the prepared: Incentive and task-informative cues interact to enhance attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 52–66. https://doi.org/10.1037/xhp0000129

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Corps, R. E., & Rabagliati, H. (2020). How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form. *Journal of Memory and Language*, *113*, 104114. https://doi.org/10.1016/j.jml.2020.104114

Costa, V. D. (2016). *More than Meets the Eye: The Relationship between Pupil Size and Locus Coeruleus Activity*.

Dambacher, M., Hübner, R., & Schlössser, J. (2011). Monetary Incentives in Speeded Perceptual Decision: Effects of Penalizing Errors Versus Slow Responses. *Frontiers in Psychology, 2*. https://www.frontiersin.org/article/10.3389/fpsyg.2011.00248

Darwin, C. (n.d.). *Praat scripts for producing Shannon AM speech*. Retrieved September 3, 2021, from http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/Shannon

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., & Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of the United States of America, 104*(41), 16032–16037. https://doi.org/10.1073/pnas.0701309104

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1–2), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(5), E618-625. https://doi.org/10.1073/pnas.1317557111

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., Sperling, R. A., Atri, A., Growdon, J. H., Hyman, B. T., Morris, J. C., Fischl, B., & Buckner, R. L. (2009). The Cortical Signature of Alzheimer’s Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and is Detectable in Asymptomatic Amyloid-Positive Individuals. *Cerebral Cortex*, *19*(3), 497–510. https://doi.org/10.1093/cercor/bhn113

Dingemanse, J. G., & Goedegebure, A. (2019). The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests. *Trends in Hearing*, *23*, 2331216519838672. https://doi.org/10.1177/2331216519838672

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. https://doi.org/10.1016/j.tics.2008.01.001

Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E., & Pouget, A. (2015). Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *ELife*, *4*, e06678. https://doi.org/10.7554/eLife.06678

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. J. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and Hearing*, *37*, 101S. https://doi.org/10.1097/AUD.0000000000000300

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., von Cramon, D. Y., & Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences*, *105*(16), 6173–6178. https://doi.org/10.1073/pnas.0708965105

Ellis, D. P. W. (2003). *Dynamic Time Warp in Matlab*. https://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/

Ellis, D. P. W. (2009). *Gammatone-like spectrograms*. https://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/

Engelmann, J. B., Damaraju, E., Padmala, S., & Pessoa, L. (2009). Combined Effects of Attention and Motivation on Visual Task Performance: Transient and Sustained

Motivational Effects. *Frontiers in Human Neuroscience*, *3*, 4. https://doi.org/10.3389/neuro.09.004.2009

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, *7*, 116. https://doi.org/10.3389/fnsys.2013.00116

Esterman, M., Reagan, A., Liu, G., Turner, C., & DeGutis, J. (2014). Reward reveals dissociable aspects of sustained attention. *Journal of Experimental Psychology: General*, *143*(6), 2287–2295. https://doi.org/10.1037/xge0000019

Feldman, H., & Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Felty, R., Buchwald, A., Gruenenfelder, T. M., & Pisoni, D. B. (2013). Misperceptions of spoken words: Data from a random sample of American English words. *The Journal of the Acoustical Society of America*, *134*(1), 572–585. https://doi.org/10.1121/1.4809540

Fong, Y., Rue, H., & Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics (Oxford, England)*, *11*(3), 397–412. https://doi.org/10.1093/biostatistics/kxp053

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gabry, J., Mahr, T., & Bürkner, P.-C. (2018). *bayesplot: Plotting for Bayesian Models*.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4). https://doi.org/10.1214/08-AOAS191

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252

Gurgel, R. K., Ward, P. D., Schwartz, S., Norton, M. C., Foster, N. L., & Tschanz, J. T. (2014). Relationship of Hearing loss and Dementia: A Prospective, Population-based Study. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *35*(5), 775–781. https://doi.org/10.1097/MAO.0000000000000313

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016. https://doi.org/10.1016/j.heares.2020.108016

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7–8), 1145–1166. https://doi.org/10.1080/01690965.2012.662280

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of*

Experimental Psychology: Human Perception and Performance, *34*(2), 460–474. https://doi.org/10.1037/0096-1523.34.2.460

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 283–295. https://doi.org/10.1037/a0020772

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 31.

Hohwy, J. (2012). Attention and Conscious Perception in the Hypothesis Testing Brain. *Frontiers in Psychology*, *3*, 96. https://doi.org/10.3389/fpsyg.2012.00096

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. https://doi.org/10.1037/a0038339

Huyck, J. J., & Johnsrude, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *The Journal of the Acoustical Society of America*, *131*(3), EL236–EL242. https://doi.org/10.1121/1.3685511

Jepma, M., & Nieuwenhuis, S. (2011). Pupil Diameter Predicts Changes in the Exploration–Exploitation Trade-off: Evidence for the Adaptive Gain Theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences*, *107*(19), 8871–8876. https://doi.org/10.1073/pnas.1002007107

Kang, O., & Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. *Consciousness and Cognition*, *35*, 128–135. https://doi.org/10.1016/j.concog.2015.05.001

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science (New York, N.Y.)*, *303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910

Knapen, T., Gee, J. W. de, Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. *PLOS ONE*, *11*(5), e0155574. https://doi.org/10.1371/journal.pone.0155574

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106–112. https://doi.org/10.1016/j.heares.2018.07.011

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, *406*, 108255. https://doi.org/10.1016/j.heares.2021.108255

Krebs, R. M., & Woldorff, M. G. (2017). Cognitive Control and Reward. In *The Wiley Handbook of Cognitive Control* (pp. 422–439). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118920497.ch24

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., Simonsick, E. M., & Health ABC Study Group. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, *173*(4), 293–299. https://doi.org/10.1001/jamainternmed.2013.1868

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science (New York, N.Y.)*, *315*(5810), 393–395. https://doi.org/10.1126/science.1131295

Mathôt, S. (2013). *A simple way to reconstruct pupil size during eye blinks*. 4.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193–203. https://doi.org/10.1111/psyp.12772

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG. *Trends in Hearing*, *21*, 2331216517706396. https://doi.org/10.1177/2331216517706396

Murphy, P. R., O’Connell, R. G., O’Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O’connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Myndex^{TM} Web Help—Luminance Contrast and Perception. (n.d.). Myndex. Retrieved August 6, 2022, from hhttps://www.myndex.com/WEB/LuminanceContrast

Naicker, P., Anoopkumar-Dukie, S., Grant, G. D., Neumann, D. L., & Kavanagh, J. J. (2016). Central cholinergic pathway involvement in the regulation of pupil diameter, blink rate and cognitive function. *Neuroscience*, *334*, 180–190. https://doi.org/10.1016/j.neuroscience.2016.08.009

Notebaert, W., & Braem, S. (2016). Parsing the effects of reward on cognitive control. In *Motivation and cognitive control* (pp. 105–122). Routledge/Taylor & Francis Group.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(3), 633–640. https://doi.org/10.1093/cercor/bhp128

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Visual Cognition*, *23*(1–2), 272–290. https://doi.org/10.1080/13506285.2014.974729

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, Cognition, and Healthy Aging: Social and Public Health Implications of the Links between Age-Related Declines in Hearing and Cognition. *Seminars in Hearing*, *36*(3), 122–139. https://doi.org/10.1055/s-0035-1555116

Plain, B., Richter, M., Zekveld, A. A., Lunner, T., Bhuiyan, T., & Kramer, S. E. (2021). Investigating the Influences of Task Demand and Reward on Cardiac Pre-Ejection Period Reactivity During a Speech-in-Noise Task. *Ear and Hearing*, *42*(3), 718–731. https://doi.org/10.1097/AUD.0000000000000971

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *13*(5), 687–697. https://doi.org/10.1162/089892901750363235

Polson, N. G., & Scott, J. G. (2012). On the Half-Cauchy Prior for a Global Scale Parameter. *Bayesian Analysis*, *7*(4), 887–902. https://doi.org/10.1214/12-BA730

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications, 7*(1), 13289. https://doi.org/10.1038/ncomms13289

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949. https://doi.org/10.1126/science.7233191

Richter, M. (2016). The Moderating Effect of Success Importance on the Relationship Between Listening Demand and Listening Effort. *Ear and Hearing*, *37*, 111S. https://doi.org/10.1097/AUD.0000000000000295

Rij, J. van, Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (2.4.1). https://CRAN.R-project.org/package=itsadug

Ritz, H., Wild, C., & Johnsrude, I. (2021). *Parametric cognitive load reveals hidden costs in the neural processing of perfectly intelligible degraded speech* (p. 2020.10.02.324509). bioRxiv. https://doi.org/10.1101/2020.10.02.324509

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, *27*(1), 33–45. https://doi.org/10.1037/a0026231

Rogers, C. S., & Wingfield, A. (2015). Stimulus-independent semantic bias misdirects word recognition in older adults. *The Journal of the Acoustical Society of America*, *138*(1), EL26-30. https://doi.org/10.1121/1.4922363

Rosemann, S., & Thiel, C. M. (2020). Neuroanatomical changes associated with age-related hearing loss and listening effort. *Brain Structure and Function*, *225*(9), 2689–2700. https://doi.org/10.1007/s00429-020-02148-w

Rysop, A. U., Schmitt, L.-M., Obleser, J., & Hartwigsen, G. (2021). Neural modelling of the semantic predictability gain under challenging listening conditions. *Human Brain Mapping*, *42*(1), 110–127. https://doi.org/10.1002/hbm.25208

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and Linguistic Interactions in Speech Perception. *NeuroImage*, *39*(3), 1444–1456. https://doi.org/10.1016/j.neuroimage.2007.09.052

Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex*, *25*(9), 2763–2773. https://doi.org/10.1093/cercor/bhu072

Shavit-Cohen, K., & Zion Golumbic, E. (2019). The Dynamics of Attention Shifts Among Concurrent Speech in a Naturalistic Multi-speaker Virtual Environment. *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnhum.2019.00386

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Signoret, C., Johnsrude, I., Classon, E., & Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 277–285. https://doi.org/10.1037/xhp0000442

Signoret, C., & Rudner, M. (2019). Hearing Impairment and Perceived Clarity of Predictable Speech. *Ear and Hearing*, *40*(5), 1140–1148. https://doi.org/10.1097/AUD.0000000000000689

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the Brain: Mind Wandering Reduces Cortical Analysis of External Events. *Journal of Cognitive Neuroscience*, *20*(3), 458–469. https://doi.org/10.1162/jocn.2008.20037

Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia*, *63*, 293–303. https://doi.org/10.1016/j.neuropsychologia.2014.09.007

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*(12), E1747–E1756. https://doi.org/10.1073/pnas.1523266113

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 186–199. https://doi.org/10.1037/a0033206

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. 48.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical activity predicts which older adults recognize speech in noise and when. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(9), 3929–3937. https://doi.org/10.1523/JNEUROSCI.2908-14.2015

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Teubner-Rhodes, S. E., Dubno, J. R., & Eckert, M. A. (2016). Cingulo-Opercular Function During Word Recognition in Noise for Older Adults with Hearing Loss. *Experimental Aging Research*, *42*(1), 67–82. https://doi.org/10.1080/0361073X.2016.1108784

Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The Cingulo-Opercular Network Provides Word-Recognition Benefit. *Journal of Neuroscience*, *33*(48), 18979–18986.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

Weissman, D. H., Gopalakrishnan, A., Hazlett, C. J., & Woldorff, M. G. (2005). Dorsal Anterior Cingulate Cortex Resolves Conflict from Distracting Stimuli by Boosting Attention toward Relevant Events. *Cerebral Cortex*, *15*(2), 229–237. https://doi.org/10.1093/cercor/bhh125

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wieling, M. (2021a). *Generalized additive modeling for EEG data*. https://www.let.rug.nl/wieling/Statistics/GAM-EEG/lab/#5_Modeling_differences

Wieling, M. (2021b). *Generalized additive models for EEG data*. http://www.let.rug.nl/wieling/Statistics/GAM-EEG/#1

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8456–8460. https://doi.org/10.1073/pnas.1201858109

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, *60*(2), 1490–1502. https://doi.org/10.1016/j.neuroimage.2012.01.035

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention.

Journal of Neuroscience, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends in Hearing*, *20*, 2331216516669723. https://doi.org/10.1177/2331216516669723

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, *36*(4), e153–e165. https://doi.org/10.1097/AUD.0000000000000145

Winn, M. B., & Moore, A. N. (2018). Pupillometry Reveals That Context Benefit in Speech Perception Can Be Disrupted by Later-Occurring Sounds, Especially in Listeners With Cochlear Implants. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518808962

Winn, M. B., & Teece, K. H. (2021). Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends in Hearing*, *25*, 23312165211027690. https://doi.org/10.1177/23312165211027688

Wood, S. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, *100*(1), 221–228.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wood, S., Wood, M. S., Yes, L., Yes, B., & Yes, N. (2013). *Package "mgcv."*

Zarei, M., Ibarretxe-Bilbao, N., Compta, Y., Hough, M., Junque, C., Bargallo, N., Tolosa, E., & Martí, M. J. (2013). Cortical thinning is associated with disease stages and dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(8), 875–882. https://doi.org/10.1136/jnnp-2012-304126

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836. https://doi.org/10.1016/j.neuroimage.2006.04.199

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 2331216518777174. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

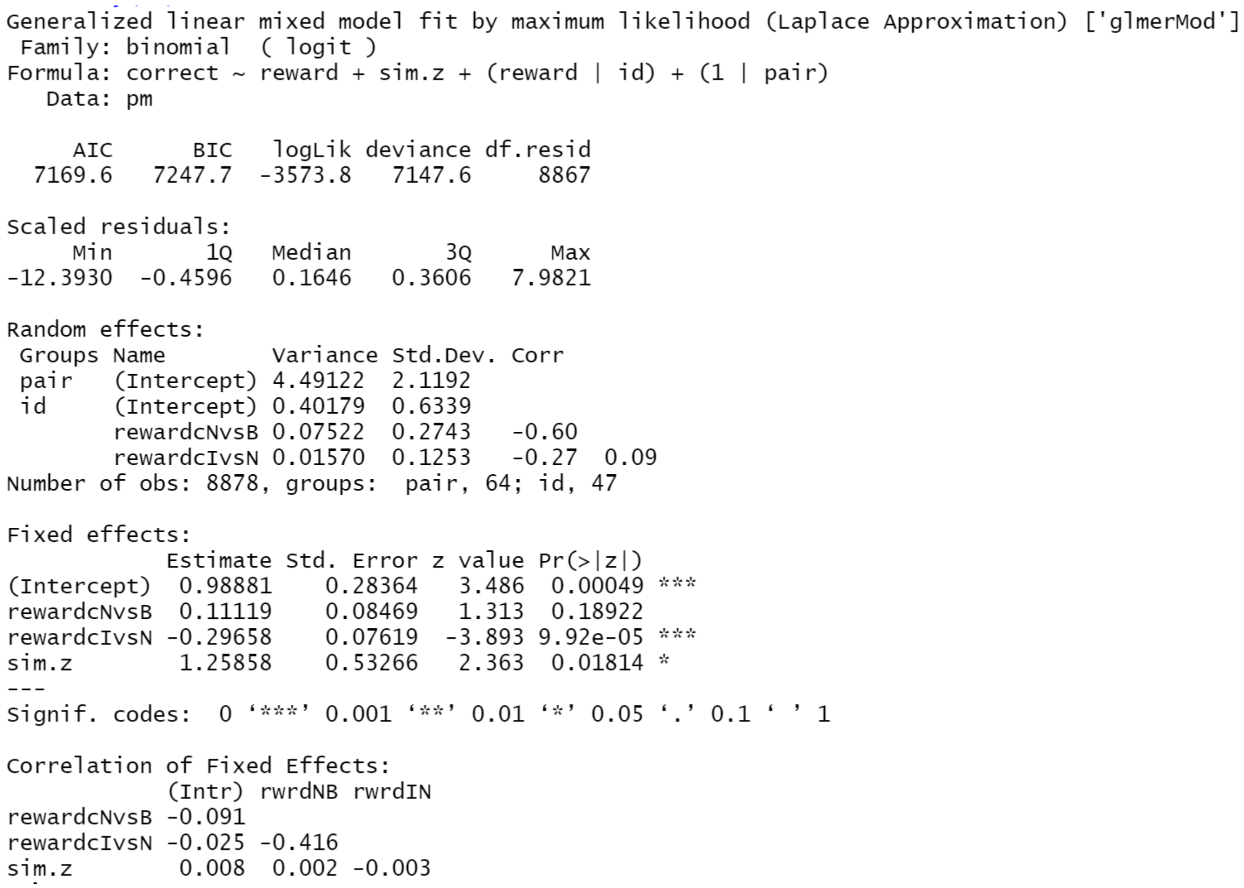
Zénon, A. (2017). Time-domain analysis for extracting fast-paced pupil responses. *Scientific Reports*, *7*(1), 41484. https://doi.org/10.1038/srep41484

Zhang, M., Siegle, G. J., McNeil, M. R., Pratt, S. R., & Palmer, C. (2019). The role of reward and task demand in value-based strategic allocation of auditory comprehension effort. *Hearing Research*, *381*, 107775. https://doi.org/10.1016/j.heares.2019.107775

Zhao, S., Bury, G., Milne, A., & Chait, M. (2019). Pupillometry as an Objective Measure of Sustained Attention in Young and Older Listeners. *Trends in Hearing*, *23*, 2331216519887815. https://doi.org/10.1177/2331216519887815

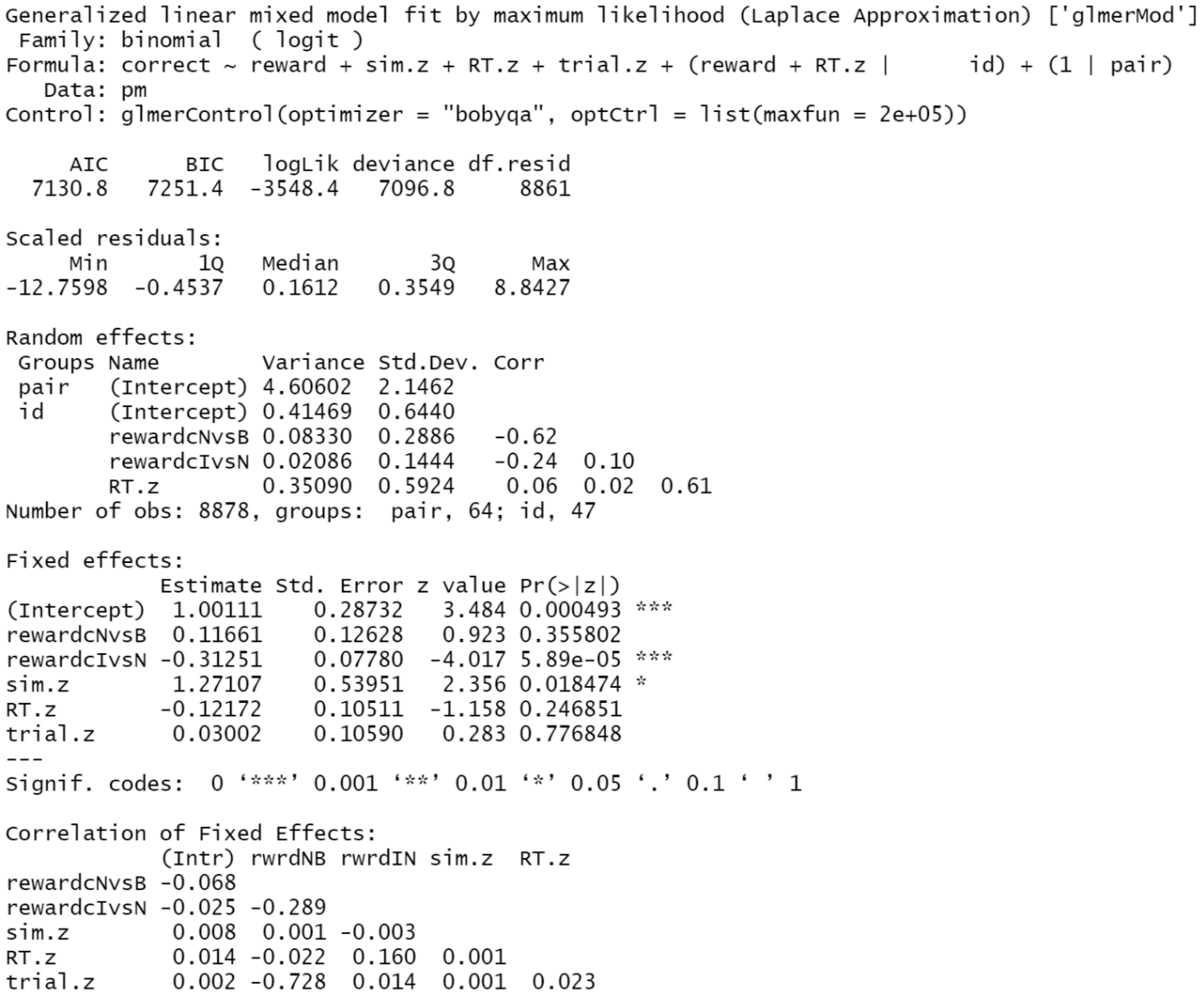
Appendices

Appendix A: Full model output for the logistic mixed model



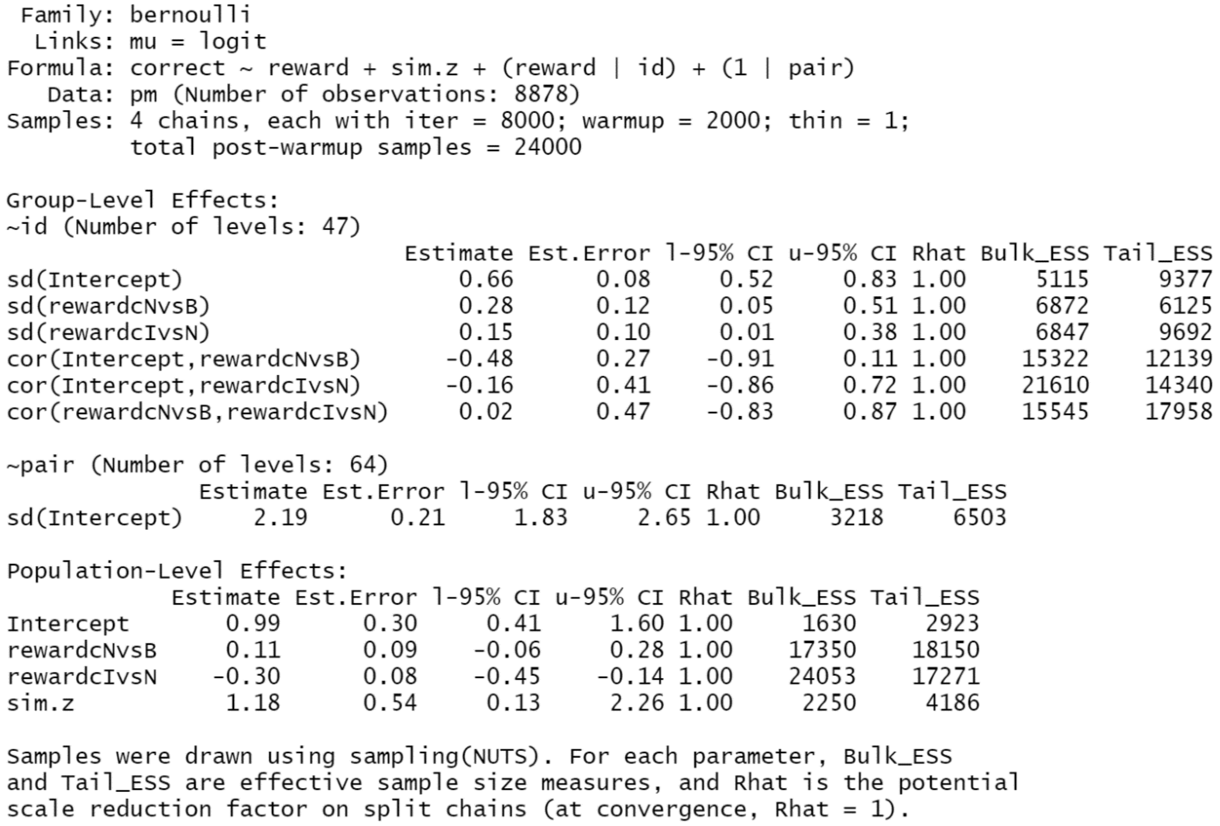
```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: correct ~ reward + sim.z + (reward | id) + (1 | pair)
   Data: pm

     AIC      BIC   logLik deviance df.resid
  7169.6   7247.7  -3573.8   7147.6     8867

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.3930  -0.4596   0.1646   0.3606   7.9821

Random effects:
 Groups Name        Variance Std.Dev. Corr
 pair   (Intercept) 4.49122  2.1192
 id     (Intercept) 0.40179  0.6339
        rewardcNvsB 0.07522  0.2743   -0.60
        rewardcIvsN 0.01570  0.1253   -0.27  0.09
Number of obs: 8878, groups:  pair, 64; id, 47

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.98881    0.28364   3.486  0.00049 ***
rewardcNvsB  0.11119    0.08469   1.313  0.18922
rewardcIvsN -0.29658    0.07619  -3.893 9.92e-05 ***
sim.z        1.25858    0.53266   2.363  0.01814 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) rwrdNB rwrdIN
rewardcNvsB -0.091
rewardcIvsN -0.025 -0.416
sim.z        0.008  0.002 -0.003
```

Appendix B: Full model output for the logistic mixed model with nuisance factors



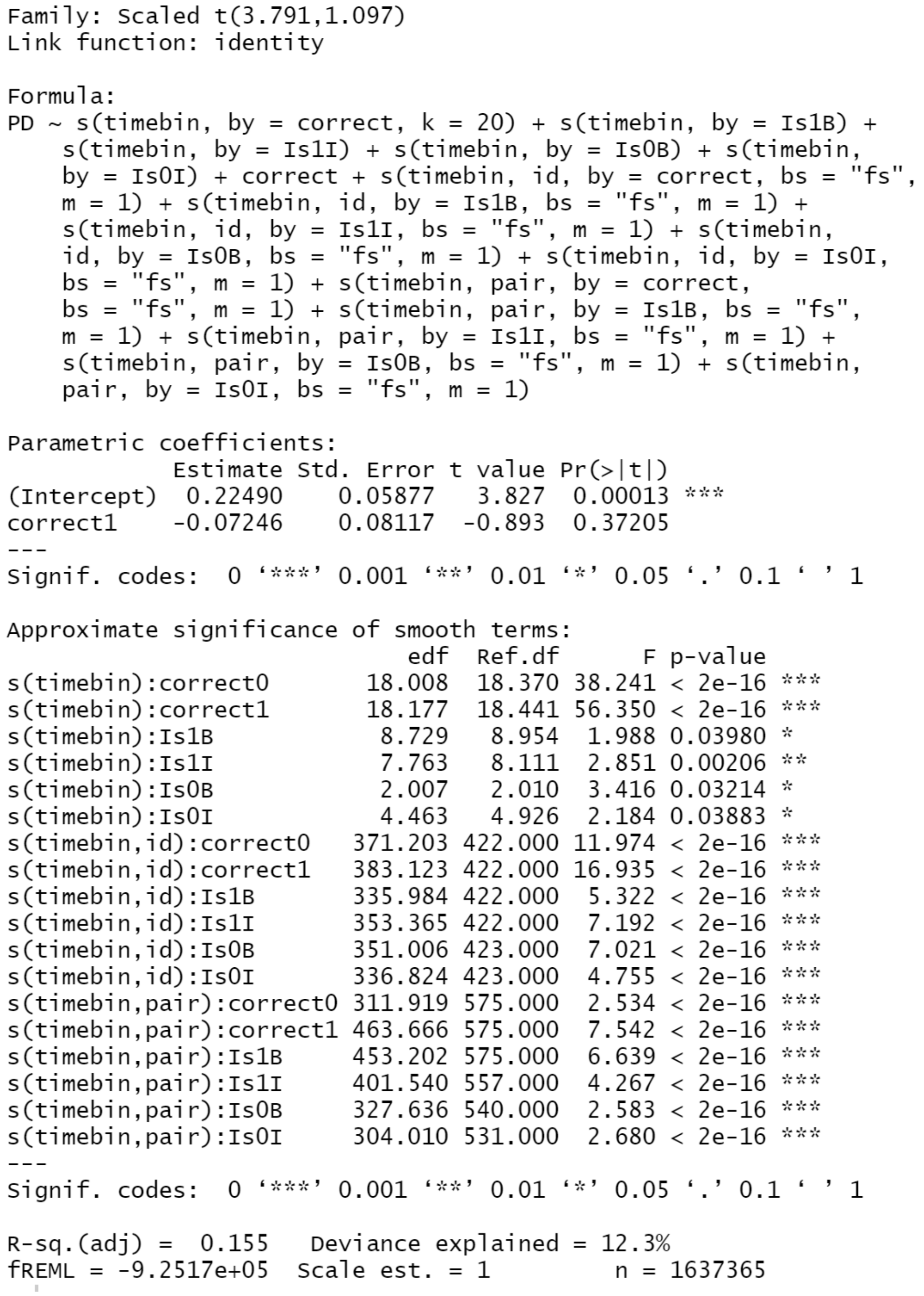
```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: correct ~ reward + sim.z + RT.z + trial.z + (reward + RT.z |      id) + (1 | pair)
   Data: pm
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))

     AIC      BIC   logLik deviance df.resid
  7130.8   7251.4  -3548.4   7096.8     8861

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.7598  -0.4537   0.1612   0.3549   8.8427

Random effects:
 Groups Name        Variance Std.Dev. Corr
 pair   (Intercept) 4.60602  2.1462
 id     (Intercept) 0.41469  0.6440
        rewardcNvsB 0.08330  0.2886   -0.62
        rewardcIvsN 0.02086  0.1444   -0.24  0.10
        RT.z        0.35090  0.5924    0.06  0.02  0.61
Number of obs: 8878, groups:  pair, 64; id, 47

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.00111    0.28732   3.484 0.000493 ***
rewardcNvsB  0.11661    0.12628   0.923 0.355802
rewardcIvsN -0.31251    0.07780  -4.017 5.89e-05 ***
sim.z        1.27107    0.53951   2.356 0.018474 *
RT.z        -0.12172    0.10511  -1.158 0.246851
trial.z      0.03002    0.10590   0.283 0.776848
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) rwrdNB rwrdIN sim.z  RT.z
rewardcNvsB -0.068
rewardcIvsN -0.025 -0.289
sim.z        0.008  0.001 -0.003
RT.z         0.014 -0.022  0.160  0.001
trial.z      0.002 -0.728  0.014  0.001  0.023
```

Appendix C: Full model output for Bayesian logistic mixed model



```
 Family: bernoulli
  Links: mu = logit
Formula: correct ~ reward + sim.z + (reward | id) + (1 | pair)
   Data: pm (Number of observations: 8878)
Samples: 4 chains, each with iter = 8000; warmup = 2000; thin = 1;
         total post-warmup samples = 24000

Group-Level Effects:
~id (Number of levels: 47)
                             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                    0.66      0.08     0.52     0.83 1.00     5115     9377
sd(rewardcNvsB)                  0.28      0.12     0.05     0.51 1.00     6872     6125
sd(rewardcIvsN)                  0.15      0.10     0.01     0.38 1.00     6847     9692
cor(Intercept,rewardcNvsB)      -0.48      0.27    -0.91     0.11 1.00    15322    12139
cor(Intercept,rewardcIvsN)      -0.16      0.41    -0.86     0.72 1.00    21610    14340
cor(rewardcNvsB,rewardcIvsN)     0.02      0.47    -0.83     0.87 1.00    15545    17958

~pair (Number of levels: 64)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.19      0.21     1.83     2.65 1.00     3218     6503

Population-Level Effects:
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept       0.99      0.30     0.41     1.60 1.00     1630     2923
rewardcNvsB     0.11      0.09    -0.06     0.28 1.00    17350    18150
rewardcIvsN    -0.30      0.08    -0.45    -0.14 1.00    24053    17271
sim.z           1.18      0.54     0.13     2.26 1.00     2250     4186

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Appendix D: Full model output for GAMM pupil dilation analysis



```
Family: Scaled t(3.791,1.097)
Link function: identity

Formula:
PD ~ s(timebin, by = correct, k = 20) + s(timebin, by = Is1B) +
    s(timebin, by = Is1I) + s(timebin, by = Is0B) + s(timebin,
    by = Is0I) + correct + s(timebin, id, by = correct, bs = "fs",
    m = 1) + s(timebin, id, by = Is1B, bs = "fs", m = 1) +
    s(timebin, id, by = Is1I, bs = "fs", m = 1) + s(timebin,
    id, by = Is0B, bs = "fs", m = 1) + s(timebin, id, by = Is0I,
    bs = "fs", m = 1) + s(timebin, pair, by = correct,
    bs = "fs", m = 1) + s(timebin, pair, by = Is1B, bs = "fs",
    m = 1) + s(timebin, pair, by = Is1I, bs = "fs", m = 1) +
    s(timebin, pair, by = Is0B, bs = "fs", m = 1) + s(timebin,
    pair, by = Is0I, bs = "fs", m = 1)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22490    0.05877   3.827  0.00013 ***
correct1    -0.07246    0.08117  -0.893  0.37205
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                             edf  Ref.df      F p-value
s(timebin):correct0       18.008  18.370 38.241 < 2e-16 ***
s(timebin):correct1       18.177  18.441 56.350 < 2e-16 ***
s(timebin):Is1B            8.729   8.954  1.988 0.03980 *
s(timebin):Is1I            7.763   8.111  2.851 0.00206 **
s(timebin):Is0B            2.007   2.010  3.416 0.03214 *
s(timebin):Is0I            4.463   4.926  2.184 0.03883 *
s(timebin,id):correct0   371.203 422.000 11.974 < 2e-16 ***
s(timebin,id):correct1   383.123 422.000 16.935 < 2e-16 ***
s(timebin,id):Is1B       335.984 422.000  5.322 < 2e-16 ***
s(timebin,id):Is1I       353.365 422.000  7.192 < 2e-16 ***
s(timebin,id):Is0B       351.006 423.000  7.021 < 2e-16 ***
s(timebin,id):Is0I       336.824 423.000  4.755 < 2e-16 ***
s(timebin,pair):correct0 311.919 575.000  2.534 < 2e-16 ***
s(timebin,pair):correct1 463.666 575.000  7.542 < 2e-16 ***
s(timebin,pair):Is1B     453.202 575.000  6.639 < 2e-16 ***
s(timebin,pair):Is1I     401.540 557.000  4.267 < 2e-16 ***
s(timebin,pair):Is0B     327.636 540.000  2.583 < 2e-16 ***
s(timebin,pair):Is0I     304.010 531.000  2.680 < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.155   Deviance explained = 12.3%
fREML = -9.2517e+05  Scale est. = 1         n = 1637365
```