Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (in press). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* DOI: 10.1037/xlm0000744

**Manuscript accepted for publication at the Journal of Experimental Psychology: Learning, Memory, and Cognition (date of acceptance: June 4th, 2019)**

# Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension

Greta Kaufeld[a]*, Anna Ravenschlag[a], Antje S. Meyer[a,b], Andrea E. Martin[a], and Hans Rutger Bosker[a]

[a]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands;* [b]*Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands*

Greta Kaufeld *(corresponding author)*; Max Planck Institute for Psycholinguistics, P.O Box 310, 6500AH Nijmegen, the Netherlands; orcid: 0000-0003-0470-442X; email: greta.kaufeld@mpi.nl; Anna Ravenschlag; Antje S. Meyer; email: antje.meyer@mpi.nl; Andrea E. Martin; orcid: 0000-0002-3395-7234; email: andrea.martin@mpi.nl; Hans Rutger Bosker; orcid: 0000-0002-2628-7738; email: hansrutger.bosker@mpi.nl

# Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension

During spoken language comprehension, listeners make use of both knowledge-based and signal-based sources of information, but little is known about how cues from these distinct levels of representational hierarchy are weighted and integrated online. In an eye-tracking experiment using the visual world paradigm, we investigated the flexible weighting and integration of morphosyntactic gender marking (a knowledge-based cue) and contextual speech rate (a signal-based cue). We observed that participants used the morphosyntactic cue immediately to make predictions about upcoming referents, even in the presence of uncertainty about the cue's reliability. Moreover, we found speech rate normalization effects in participants' gaze patterns even in the presence of preceding morphosyntactic information. These results demonstrate that cues are weighted and integrated flexibly online, rather than adhering to a strict hierarchy. We further found rate normalization effects in the looking behavior of participants who showed a strong behavioral preference for the morphosyntactic gender cue. This indicates that rate normalization effects are robust and potentially automatic. We discuss these results in light of theories of cue integration and the two-stage model of acoustic context effects.

Keywords: language comprehension; speech perception; cue integration; rate normalization

When comprehending spoken language, listeners make use of multiple cues from different information sources and across several hierarchical levels of linguistic representations. A distinction is commonly made between cues from at least two sources: acoustic, or "signal-based" cues, and linguistic, or "knowledge-based" cues. Signal-based cues include the spectral and temporal properties of the acoustic speech signal, such as voice onset time (VOT; e.g., Lisker & Abramson, 1967; Toscano & McMurray, 2015) and contextual speech rate (e.g., Bosker, 2017; Maslowski, Meyer, & Bosker, 2018; Reinisch & Sjerps, 2013). Knowledge-based cues, on the other hand, include knowledge about phonotactic and syntactic constraints (e.g., Huettig & Janse, 2016; McQueen, 1998; Tuinman, Mitterer, & Cutler, 2014), as well as semantic context (Altmann & Kamide, 1999; Wicha, Moreno, & Kutas, 2004). Consequently, many models of spoken word and language comprehension incorporate at least some degree of interaction between information from both knowledge-based and signal-based information sources (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986), but few of them make predictions about how the brain computationally integrates this available information from different levels of linguistic hierarchy (although see, e.g., Norris & McQueen, 2008, for a Bayesian implementation of lexical recognition).

The goal of the current study is to contribute to our understanding of language comprehension by investigating how signal-based and knowledge-based cues are integrated and weighted against each other during online speech comprehension. Using eye-tracking within the visual world paradigm, we investigate two questions: 1) Are knowledge-based, morphosyntactic cues towards grammatical gender immediately used to generate predictions about upcoming referents, even in the presence of uncertainty? 2) Are signal-based, contextual speech rate cues used even in the presence of preceding morphosyntactic information? We also investigate, for the first time, variations in the

strategies that participants employ when integrating cues with each other by mapping participants' behavioral responses to their eye-tracking data. We discuss the implications of our findings within the framework of cue integration (Martin, 2016) and the two-stage model of acoustic context effects (Bosker et al., 2017).

**Language processing as hierarchical cue integration**

Drawing on principles from perception, speech processing, and neurophysiology, Martin (2016) suggested a framework of cue integration for language processing, offering a general mechanism of how the brain utilizes cues across multiple levels of hierarchy to comprehend and produce language (see, e.g., Ernst & Bülthoff, 2004; Fetsch, DeAngelis, & Angelaki, 2013, for detailed descriptions of cue integration for visual and multisensory perception). Within cue integration frameworks, relevant cues are *combined* by means of summation and *integrated* by normalization against all other available cues. Each cue has an associated *weight*, which is a formalization of how reliable the cue is in a given situation and in combination with all other cues. Cue weights can be dynamically updated, which gives the system the flexibility to generate robust percepts even in the presence of uncertainty, noise and variability.

Models related to cue integration have previously been suggested for phoneme categorization (e.g., McMurray & Jongman, 2011) and lexical recognition (e.g., Norris & McQueen, 2008). Martin (2016) suggested a cascading cue integration architecture across all levels of language processing, where functional equivalents of formal linguistic representations can emerge from sensory cues, and can in turn act as cues for higher-level representations. For speech comprehension, this means extracting and integrating relevant cues from signal-based and knowledge-based sources in order to infer higher-level linguistic information and meaning (Martin, 2016).

Establishing a hierarchical inventory of cues for spoken language comprehension remains a challenging objective for psycholinguistic research. Based on a series of experiments in which the amount and reliability of information from cues at different levels of representation was systematically manipulated, Mattys, White, and Melhorn (2005) proposed a hierarchically organized model of lexical segmentation. According to the original version of their model, cues are organized into three hierarchical tiers consisting of lexical (Tier I), segmental (Tier II), and metrical prosodic (Tier III) cues. Crucially, cues from Tier I, which can include contextual, syntactic, semantic, and morphological information, form the highest level of the hierarchy and can override cues from the lower two levels of representation (Mattys et al., 2005). However, in a subsequent set of experiments, Mattys, Melhorn, and White (2007) found that effects of syntactic knowledge on lexical segmentation could be attenuated and modulated by conflicting acoustic cues. Using a word monitoring task, they assessed how participants processed the combination of a morphosyntactic cue (singular vs. plural lexical information; e.g., *those women* vs. *that woman*) with a subsequent acoustic cue (pivotal /s/, e.g., *take#spins* vs. *takes#pins*). In a neutral listening situation without preceding syntactic information, listeners made use of acoustic cues for segmentation, as evidenced by faster target detection times for *pins* in *takes#pins*, and *spins* in *take#spins*. When preceded by a plural noun phrase, the syntactic cue took precedence over the acoustic cue (i.e., faster target detection for *spins* in "*those women take#spins"* and "*those women takes#pins"*). This result is in line with a hierarchical model of speech processing, where syntactic cues can "override" acoustic cues. For singular noun phrases, however, no effect of superiority for the syntactic cue was found, showing the same pattern of results as for the neutral condition (i.e., faster target detection for *pins* in *"that woman takes#pins"*, and *spins* in *"that woman take#spins*"). Mattys et al. (2007)

therefore proposed a graded, dynamic relationship between knowledge-based and signal-based cues. The concept of a dynamic link between cues from different levels of hierarchy, although not mathematically formalized in the model by Mattys and colleagues (2007), bears striking similarities to cue weighting and normalization as suggested by linguistic models of cue integration (Martin, 2016).

**Integrating and weighting knowledge- and signal-based cues**

A growing body of research has investigated the interplay between signal-based and knowledge-based cues. Most relevant for our purposes are studies investigating contextual speech rate cues. The speech rate in a lead-in sentence can change the perception of a following target word: For instance, a vowel ambiguous between short /ɑ/ and long /a:/ in Dutch is perceived as /a:/ in the context of a fast speech rate because it sounds relatively long compared to the short vowels in the fast context, but as /ɑ/ in the context of a slow speech rate (Bosker, 2017; Bosker & Reinisch, 2017; Maslowski et al., 2018; Reinisch & Sjerps, 2013). This process, known as rate normalization, influences many duration-cued phonemic contrasts, such as singleton-geminate (Mitterer, 2018), /b/-/p/ (Gordon, 1988), /b/-/w/ (Wade & Holt, 2005), and recognition of unstressed syllables (*form* vs. *forum*; Baese-Berk, Dilley, Henry, Vinke, & Banzina, 2018) and words (*silver jewelry* vs. *silver or jewelry*; Dilley & Pitt, 2010; *cease* vs. *see us*; Baese-Berk et al., 2018). Importantly, contextual rate effects have been shown to arise very rapidly during spoken word comprehension, and have thus been hypothesized to occur at the earliest stages of perception (e.g., Bosker & Ghitza, 2018; Toscano & McMurray, 2015; Reinisch & Sjerps, 2013).

How exactly contextual speech rate cues interact with knowledge-based cues during online speech processing is unclear. For example, Morrill, Baese-Berk, Heffner, and Dilley (2015) examined the interacting effects of contextual speech rate and

linguistic knowledge on reduced word recognition using a transcription task. They presented participants with utterances that included highly reduced function words, such as "or" in the sentence "Don must see the harbor [or] boats". Depending on the perception of the reduced function word "or" (in square brackets), this sentence could be interpreted as either "Don must see the harbor boats" or "Don must see the harbor or boats". Crucially, the rate of the surrounding context (underlined in the example) was manipulated to be either slowed or unaltered. Morrill et al. (2015) observed that slowing down the speech rate in the context made the reduced function word "or" perceptually disappear: Participants transcribed the sentence without the critical function word (e.g., "harbor boats" rather than "harbor or boats"). Moreover, even when the reduced function word was syntactically obligatory (e.g., "Conner knew that bread and butter [are] both in the pantry", where the sentence is only grammatical if the function word "are" is perceived as being present), participants still transcribed the sentence without the function word if it was embedded in slow speech. In fact, the effect of contextual speech rate was even observed to be comparable across syntactically optional and syntactically obligatory sentences, and no significant interaction was found between speech rate and syntactic obligatoriness, suggesting that the weighting of contextual speech rate was not modulated by conflicting syntactic cues. Contrasting older and younger speakers, Heffner, Newman, Dilley, and Idsardi (2015) reported similar results: Presented with similar stimuli as used in Morrill et al. (2015), participants in both age groups were less likely to report a critical word if it was 1) presented in a slow context, and 2) syntactically optional. Again, the interaction between the two predictors was non-significant, suggesting that participants made use of knowledge-based and signal-based cues independently.

The observation in Morrill et al. (2015) and Heffner et al. (2015) that the weighting of contextual speech rate as a cue to lexical recognition is not modulated by conflicting syntactic cues seems to clash with Mattys et al.'s (2005) proposal that syntactic knowledge operates at the highest tier of lexical recognition. At the same time, the findings raise several questions. First, Morrill et al. (2015) and Heffner et al. (2015) used a transcription task, where participants were asked to transcribe the auditory stimuli after having heard the entire utterance. The results therefore reflect participants' explicit decision-making about the nature of the stimuli and do not offer direct insights into *when* during comprehension signal- and knowledge-based cues are extracted, combined, and weighted. Second, the critical target region in Morrill et al.'s (2015) and Heffner et al.'s (2015) stimuli always preceded the syntactic cue. That is, in a sentence like "Conner knew that bread and butter [are] both in the pantry", where perception of the word "are" was obligatory for the sentence's grammaticality, participants only discovered that the verb was syntactically obligatory *after* presentation of the critical region. This is especially interesting because Mattys et al. (2007) suggested that the time-course of knowledge-based and signal-based cues might play a crucial role in the way in which these two sources of information are integrated. If that is the case, it is possible that the absence of an interaction between acoustic and syntactic cues in Morrill et al. (2015) and Heffner et al. (2015) was due to their order in the stimuli. Finally, Morrill et al. (2015) and Heffner et al. (2015) reported group averages, but they did not investigate individual variation in cue weighting. Assuming a relative degree of flexibility in cue weighting as suggested by Martin (2016) and Mattys et al. (2007), as well as the results reported by Morrill et al. (2015) and Heffner et al. (2015), the question emerges whether individual participants also employed different strategies

during the experiment, or whether cue-weighting effects generally arise on a group level.

**Current study**

In the current experiment, we aimed to examine the flexible interplay between signal- and knowledge-based cues during online spoken language comprehension. More specifically, we used eye-tracking within the visual world paradigm to test the robustness of signal-based contextual rate cues in the presence of earlier knowledge-based cues to grammatical gender. This allowed us to investigate how the system integrates potentially conflicting cues from different levels of linguistic hierarchy. We manipulated minimal word pairs in Dutch to contain vowel tokens that were ambiguous between short /ɑ/ and long /a:/ (e.g., *vat*$_{NEUTER}$ "barrel", *vaat*$_{COMMON}$ "dishes"), embedded in carrier sentences at slow or fast speech rates. Participants were presented with two pictures on a screen, corresponding to the short /ɑ/ or long /a:/ noun (e.g., a picture of a barrel and a picture of dishes), while listening to auditory instructions at fast or slow rates asking them to look at one of the two pictures (e.g., *Kijk nu eens naar de*$_{COMMON}$*/het*$_{NEUTER}$ *ontzettend vuile vat*$_{COMMON}$*/vaat*$_{NEUTER}$*, alsjeblieft,* "Now look once at the$_{COMMON/NEUTER}$ terribly dirty barrel$_{COMMON}$/dishes$_{NEUTER}$, please). Participants then clicked on the picture which they thought corresponded to the target. Crucially, the carrier sentences contained a preceding morphosyntactic cue in the form of the definite article *de*$_{COMMON}$ or *het*$_{NEUTER}$, which has previously been shown to elicit anticipatory language processing within the visual world paradigm (Huettig & Janse, 2016). However, Huettig and Janse (2016) found that only about half of the participants showed the expected anticipatory looking behavior (see their Figure 5) in an experimental paradigm that only targeted morphosyntactic prediction. Similarly, using the same experimental paradigm, Huettig and Guerra (2019) reported evidence for

anticipatory language processing being attenuated by factors such as shorter preview time, implicit vs. explicit participant instructions, and faster speech rate of the carrier sentence. As such, it remains unclear whether morphosyntactically driven anticipatory looking behavior can be observed in an experiment with additional signal-based cues to target perception. In our experimental manipulation, the article could act as an early cue towards grammatical gender, and thus bias participants' perception towards one of the two vowel interpretations. There were thus two cues towards the "target" picture in our experiment: 1) the gender of the article preceding the noun, and 2) the contextual speech rate and its consequences for the relative perception of the temporal properties of the ambiguous vowel. Which of the two nouns participants considered the "target" was entirely up to them, depending on whether they preferred the information conveyed by the knowledge-based gender cue or the signal-based speech rate cue.

The cue integration model predicts that the system rapidly extracts and integrates signal-based and knowledge-based cues during spoken language comprehension. Our experimental manipulation allowed us to investigate the relative contribution of these cues from distinct levels of linguistic representation as a function of participants' looking preferences as the information in the sentence unfolded. We hypothesized that listeners would rapidly use the morphosyntactic gender cue conveyed by the article in order to make predictions about the ambiguous noun, which would be reflected in participants looking more toward the picture corresponding to the gender of the article. Crucially, this would occur well before the onset of the noun. The rate manipulation introduced a potential mismatch between the gender of the article and the gender of the (perceived) noun, making both cues somewhat unreliable for participants. Analyzing a time window immediately after the offset of the article, but before the onset of the ambiguous vowel, thus allowed us to address our first research question: Are

knowledge-based, morphosyntactic cues towards grammatical gender immediately used to generate predictions about upcoming referents, even in the presence of uncertainty?

Second, we asked whether listeners would take signal-based contextual rate cues into account even in the presence of preceding disambiguating, potentially conflicting, articles. This would be reflected in participants shifting their gaze toward the picture corresponding to the vowel perception elicited by the rate manipulation after hearing the ambiguous vowel. Specifically, when embedded in a slow context sentence, the ambiguous vowel should appear relatively short in contrast to the preceding speech sounds, thus biasing participants towards perceiving the vowel as /ɑ/ and looking at the corresponding picture. Conversely, fast context rates should bias participants towards looking more at the picture corresponding to an /a:/ vowel interpretation (Reinisch & Sjerps, 2013). Analyzing a time window immediately after the offset of the ambiguous vowel (i.e., the earliest moment in time when participants could access the duration cues on the vowel) until the end of the utterance thus allowed us to answer our second research question: Are signal-based, contextual speech rate cues used even in the presence of preceding, potentially conflicting, morphosyntactic information?

Regarding both of our questions, it is possible that one of the two cues is entirely overwritten by the other, and that participants base their choice of response only on the cue that they perceive as more reliable. For example, the signal-based, speech rate induced cue might be entirely overwritten by the preceding knowledge-based, morphosyntactic cue. If, as Mattys et al. (2005) suggested, syntactic cues are generally weighted more strongly than acoustic cues, we should thus not find significant changes to eye fixations as a function of contextual speech rate during the vowel window, because participants would simply weigh the syntactic cue more heavily and ignore the contextual rate manipulation. Observing more looks towards the picture corresponding

to the gender of the article in the carrier sentence, but no effect of the speech rate manipulation during the vowel window, would thus be in line with Mattys et al.'s (2005) original model. Conversely, observing rate normalization effects in the noun window, *even in the presence* of preceding morphosyntactic information, would be in line with the later model suggested by Mattys and colleagues (2007), and with more general, computationally formalized models of cue integration (Martin, 2016).

Using eye-tracking within the visual world paradigm allowed us to investigate this potentially flexible weighting of signal-based and knowledge-based cues while participants were processing the sentences online. However, it is possible that individual participants employ different strategies during cue integration and sentence comprehension (cf. Van Bergen & Bosker, 2018). If, for instance, half of our participants weighed the knowledge-based cue more heavily, while the other half relied more strongly on the signal-based cue, then standard analysis of average behavior across all participants would not be very insightful. Therefore, we also mapped participants' offline behavioral responses (target categorization mouse-clicks) to their online cue weighting behavior as evidenced in their gaze patterns. Specifically, we created a measure of each individual's preference for the knowledge-based vs. the signal-based cue based on their categorization responses, which we then linked to participants' eye fixations in the vowel window. Rather than drawing conclusions about each cue's relative weight based solely on average behavioral measures (e.g., Heffner et al., 2015; Morrill et al., 2015), we were thus able to investigate whether individual strategies were reflected in different eye-tracking patterns.

Moreover, mapping participants' behavioral responses to their eye-tracking data also allowed us to test whether we could find online evidence for rate manipulation effects in the eye fixation data for participants whose behavioral responses principally

followed the knowledge-based, syntactic cue. This question is relevant in light of debate about the robustness of phoneme-level rate effects, which some have proposed to be "fragile" (Baese-Berk et al., 2018), while others have argued that they are robust and potentially automatic (Bosker et al., 2017; Reinisch & Sjerps, 2013). Observing phoneme-level rate effects even for participants who behaviorally favored the preceding morphosyntactic cue would be strong evidence for rate normalization effects arising very early during perception, unmodulated by other information sources.

## Methods

We aimed to test 1) whether knowledge-based, morphosyntactic cues towards grammatical gender were immediately used to generate predictions about upcoming referents, and 2) whether signal-based, contextual speech rate cues persisted even in the presence of preceding, potentially conflicting, morphosyntactic information. We used eye-tracking (visual world paradigm) in order to obtain online measures of the influence of these knowledge-based and signal-based cues on the perception of the phonemic vowel contrast /ɑ/ vs. /a:/ in Dutch. We further mapped online eye-tracking data to offline behavioral data in order to investigate the strategies that individuals employ while combining different sources of information.

### Participants

Native speakers of Dutch ($N$ = 36, 19 females, $M_{age}$ = 22 years) with self-reported normal hearing were recruited from the MPI participant pool, with informed consent as approved by the Ethics Committee of the Social Sciences Department of Radboud University (Project Code: ECSW2014-1003-196). Participants were paid for their participation.

**Materials and Design**

Stimuli consisted of 7 Dutch sentences, each containing a unique /ɑ/-/a:/ minimal pair that differed in grammatical gender (common vs. neuter; e.g., *(de) as*$_{COMMON}$ "ash" - *(het) aas*$_{NEUTER}$ "bait"). Each sentence followed a specific sentence frame, for instance, *Kijk nu eens naar [de|het] ontzettend vieze [as|aas] alsjeblieft;* "Look now once to [the$_{COMMON}$|the$_{NEUTER}$] very dirty [ash$_{COMMON}$|bait$_{NEUTER}$] please" (see Table A, Supplementary Materials, for a complete list of stimuli). We recorded a female native speaker of Dutch, who was naïve to the purpose of the experiment, reading the sentences in two syntactic conditions (with *de* and *het*) and with both nouns. Recordings were made in a sound-attenuated booth and digitally sampled at 44,100 Hz on a computer located outside the booth with Audacity software (Audacity Team).

For the various speech rate and syntactic conditions, we manipulated, using PSOLA in Praat (Boersma & Weenink, 2012), the speech rate of the lead-in fragment (*Kijk nu eens naar*), adjectival phrase (e.g., *ontzettend vieze*) and the final fragment (*alsjeblieft*) of each sentence in a combined fashion through linear compression with a factor of 0.66 (fast condition) and 1.5 (1/0.66; slow condition) of the original recording. We created two syntactic conditions of each sentence by replacing the article *het* from each sentence with the article *de* from a recording of the same sentence.

For the nouns, we required vowels that were both spectrally and durationally ambiguous. The /ɑ/-/a:/ vowel contrast in Dutch is cued by both spectral (lower formant values for /ɑ/, higher formant values for /a:/) and temporal cues (shorter duration for /ɑ/, longer duration for /a:/) (Escudero, Benders, & Lipski, 2009). We created two-dimensional spectral and durational vowel continua for each vowel by first creating a linear 9-point duration continuum (1 = original duration of /ɑ/; 9 = original duration of /a:/; in steps of 12.5% of the duration difference; using PSOLA in Praat). Then, for each

duration step, we used sample-by-sample linear interpolation (9-point continuum; 1 = 100% /ɑ/ + 0% /a:/; 5 = 50% /ɑ/ + 50% /a:/; 9 = 0% /ɑ/ + 100% /a:/) to create different spectral versions of the durationally matched vowels (i.e., changing vowel quality).

We then conducted a pretest in order to choose the most suitable (i.e., the most ambiguous) combinations of duration and interpolation steps for each item pair. Participants who were naïve to the purpose of the experiment and did not participate in the main experiment ($N$ = 20, 15 females, $M_{age}$ = 23.8 years) listened to short excerpts of the created stimuli, consisting of only the adjectival phrase, noun, and outro, thus avoiding any biasing information from the article (e.g., *ontzettend vieze* $as_{COM}$/$aas_{NEU}$ *alsjeblieft*). They indicated via button press whether they had heard the word corresponding to the vowel /ɑ/ or /a:/ (e.g., *as* or *aas*). Based on the results of the pretest, we selected a unique set of five different duration steps from one and the same interpolation step for each item pair. These five steps spanned a perceptual range of relatively few long /a:/ responses (mean long /a:/ categorization of step 1 = 22%) to relatively many long /a:/ responses (mean long /a:/ categorization of step 5 = 65%). This resulted in seven unique 5-step duration continua with fixed vowel qualities.

The resulting 140 stimuli (2 syntactic conditions x 2 speech rates x 7 pairs x 5 continuum steps) formed an experimental block. Participants were presented with two blocks in an experimental session, so that each participant was exposed to 280 sentences in total. The pictures for the visual-world paradigm were selected from the MultiPic database (Duñabeitia et al., 2018) if available, or retrieved from copyright-free online resources. All pictures were scaled to a dimension of 300 pixels at the longest side.

**Procedure**

Participants were tested individually in a sound-conditioned booth. They were seated at a distance of approximately 60 cm in front of a 50,8 cm by 28,6 cm screen with a tower-

mounted Eyelink 1000 eye-tracking system (SR Research) and listened to stimuli at a comfortable volume through headphones. Stimuli were delivered using Experiment Builder software (SR Research). Eye movements were recorded using right pupil-tracking at a rate of 1000 Hz.

Each trial started with a blue fixation rectangle in the middle of the screen to center the mouse and the participant's gaze position. The rectangle disappeared when the participant clicked on it. The fixation screen was immediately followed by the presentation of the visual stimuli. After a 1 s preview interval, the auditory stimulus was played. Participants were instructed to listen to the complete auditory stimulus (no response possible before audio offset) and to click on the corresponding picture on the screen. We did not instruct participants about the (non-)grammaticality of some article plus vowel combinations (i.e., hearing $het_{NEU}$ in combination with $as_{COM}$ with a short vowel sounds ungrammatical). Thus, participants were free to choose whichever cue to base their categorization responses on. The trial ended when participants had clicked on a picture. The positioning of the visual stimuli, centered in the left or right half of the screen, was counterbalanced across participants, and the order of trials within a block was randomized across participants.

In order to get familiarized with the task, participants completed four practice trials before the experiment. After half of the experiment, participants were allowed to take a self-paced break. Including instructions, calibration and debriefing, the experimental procedure took approximately 50 to 60 minutes to complete.

## Results

### Behavioural categorization data

Due to the nature of our stimuli and the morphosyntactic regularities of the Dutch

language, we could not simply include the gender of the definite article in our statistical analyses, because there is no 1:1-mapping between each noun's gender and its associated vowel length. In other words, hearing the article *het*$_{NEUTER}$ might bias participants towards looking at the picture corresponding to a long vowel for some item pairs (e.g., *aas*$_{NEUTER}$ "bait" and *as*$_{COMMON}$ "ash"), but to a short vowel for others (e.g., *vaat*$_{COMMON}$ "dishes" and *vat*$_{NEUTER}$ "barrel"). In order to capture this variability for further statistical analyses, we decided to include a binomial Article Bias variable in our statistical analyses. To further illustrate, when presented with two pictures (e.g., *aas*$_{NEUTER}$ "bait" and *as*$_{COMMON}$ "ash"), hearing the article *het*$_{NEUTER}$ would be a cue towards a *long* vowel interpretation, whereas hearing the article *de*$_{COMMON}$ would be a cue towards a *short* interpretation. Depending on the trial-specific combination of article and pictures, each trial can thus be considered to introduce a *long* or *short* Article Bias.

Figure 1 shows participants' categorization responses (calculated as the proportion of long responses) split by the two Article Biases, collapsed across all nouns, for slow and fast context rates.

We used GLMMs with a logistic link function (Jaeger, 2008) as implemented in the lme4 library (Bates, Maechler, & Bolker, 2012) in R (R Development Core Team, 2012) in order to evaluate the binomial response corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of Article Bias (categorical predictor with two levels: article bias towards long vowel coded as +0.5; short as -0.5), Continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), and Rate (categorical predictor with two levels: fast coded as +0.5; slow as -0.5), and all their interactions. The random effects structure contained random intercepts for

Participants and Items, because adding additional random slopes resulted in non-convergence.

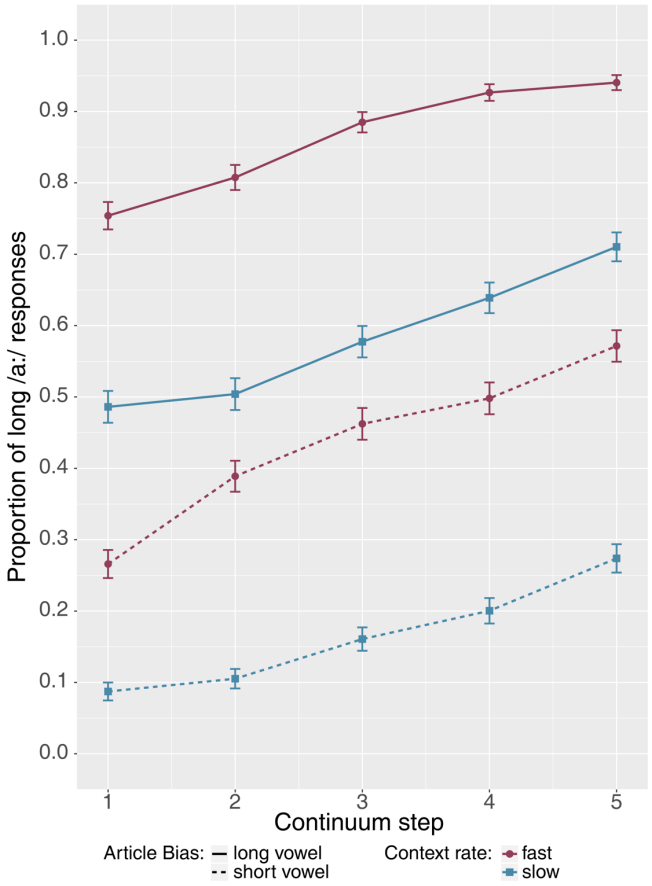


*Figure 1. Categorization curves showing the proportion of long vowel responses.* Long vowel responses are plotted as a function of duration continuum step, split for the two speech rates (red: fast rate; blue: slow rate) and the two Article Biases (solid: article biases towards 'long vowel' interpretation; dashed: article biases towards 'short vowel' interpretation). Error bars represent the standard error of the mean.

The complete model outputs are summarized in Table 1. The model revealed a significant effect of Continuum ($p < 0.001$), indicating that participants were more likely to select the "long vowel" picture at higher continuum steps. This shows that our experimental vowel manipulation was successful. The model also revealed a significant effect of Article Bias ($p < 0.001$), indicating that participants were more likely to select the "long vowel" in trials in which the preceding article was congruent with the picture corresponding to a long vowel interpretation.

Crucially, there was also a significant effect of Rate ($p < 0.001$), indicating that participants were more likely to respond with the picture corresponding to the long

vowel in fast contexts. This indicates that rate effects occurred even in the presence of earlier morphosyntactic information.

We also found a significant three-way interaction between Rate, Continuum and Article Bias ($p = 0.002$), indicating that the effect of Article Bias was slightly more pronounced at higher duration continuum steps in fast contexts. [1]

**Investigating attenuating effects of the morphosyntactic cue**

If rate effects are easily modulated and overridden by higher-level information, it is possible that the Rate effects we observe here are attenuated by the presence of the earlier morphosyntactic cue and thus smaller than they would be in isolation. We investigated this question by comparing the behavioural responses from the experiment to those from the pretest, where participants heard the manipulated vowel embedded in a fast or slow context, but without any additional morphosyntactic information (i.e., sentence excerpts excluding the article; see section "Materials and Design").

A GLMM tested the binomial responses corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of Continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), Rate (categorical predictor with two levels: fast coded as +0.5; slow as -0.5), and Experiment (categorical predictor with two levels: main experiment coded as +0.5; pretest as -0.5), and the interaction between Rate and Experiment. The model contained random intercepts for Participants and Items.

The complete model outputs are summarized in Table 2. The model revealed significant main effects of Rate ($p < 0.001$) and Continuum ($p < 0.001$), but no main

[1] In order to investigate whether these effects changed as a function of experimental block, we also tested a model including an additional fixed effect of Block and all possible interactions. This model revealed no main effect of Block and no interactions with Block.

effect of Experiment ($p = 0.337$). No interaction between Rate and Experiment was observed ($p = 0.250$), indicating that the Rate effect was not attenuated by the presence of preceding morphosyntactic information in the main experiment. We take this to suggest that the Rate effect was robust against modulation by higher-level information.

## Eye-tracking data

Prior to the analyses, blinks and saccades were excluded from the data. We divided the screen into two sections (left and right) and coded fixations on either half as a look toward that particular picture. The eye fixation data were down-sampled to 100 Hz for simplicity. We used GLMMs with a logistic link function (Jaeger, 2008) as implemented in the lme4 library (Bates et al., 2012) in R (R Development Core Team, 2012) in order to evaluate participants' eye fixations as the meaning of each sentence unfolded across time.

### Article window analysis

In order to investigate our first question, we analyzed a time window spanning from 200 ms after article onset, accounting for the time it takes to launch a saccade (Matin, Shao, & Boff, 1993) until the onset of the ambiguous word. This allowed us to test whether anticipatory language processing based on the gender information carried in the article occurred even when the article was not a univocally reliable cue towards the noun. We expected to find that items containing an article that biased towards the object corresponding to a long vowel interpretation would elicit more looks to the long object, well before the onset of the noun. Figure 2 shows participants' eye movements (calculated as the proportion of looks to the pictures of long vowel interpretation) split by Article Bias (to either the long or the short vowel interpretation) and Rate (fast vs. slow) in the article window, with the analysis window shaded in grey. Note that we do

not illustrate the article analysis window in its entirety here, because the onset of the ambiguous noun was earlier in the fast compared to the slow speech rate condition and the length of the analysis window thus differed between the two rate conditions.

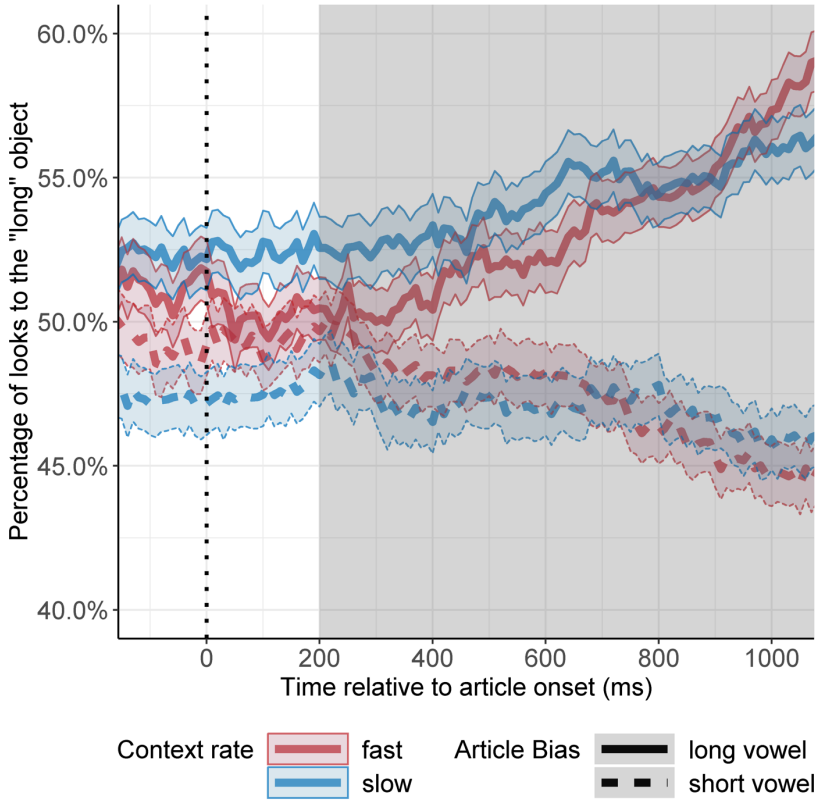


*Figure 2. Percentage of looks to the "long vowel" object across time in the article window.* Time point 0 marks the onset of the article. Looks towards the long item are plotted across time following the onset of the article in two syntactic categories (trials with a 'long vowel' Article Bias contained an article that biased participants towards the object corresponding to the long interpretation: solid line; trials with a 'short vowel' Article Bias contained an article that biased participants towards the short interpretation: dashed line) when embedded in contexts of distinct rates (fast rate: red line, slow rate: blue line). The area shaded in grey indicates an illustration of the window analysed in the article window analysis, spanning from 200 ms after article onset until the onset of the ambiguous word. Red and blue shading indicates standard error of the mean.

A GLMM tested the binomial looks to the object corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of Article Bias (categorical predictor with two levels: long coded as +0.5; short as -0.5), Time (z-scored around the mean of the analysis window), and their interaction. The random effects structure contained random intercepts for Participants and Items and by-participant and by-item random slopes for both fixed factors and their interactions. Note that we did not include Rate as

a predictor in this model because participants had not yet heard the ambiguous vowel at this point in time.

The model (see Table 3) revealed a significant effect of Article Bias ($p < 0.001$), indicating that participants were more likely to look at the picture corresponding to a long vowel interpretation if the article corresponded to that interpretation. We also found a significant interaction between Article Bias and Time, indicating that the effect of Article Bias grew over time (i.e., we observed a larger effect in later parts of the time window; $p = 0.001$).

**Vowel window analysis**

In order to investigate whether the effects of the rate manipulation on eye fixations could still be observed after the presentation of preceding morphosyntactic information (in our case, the article encoding the gender of the noun), we analyzed a vowel window ranging from 200 ms after the offset of the manipulated vowel until speech offset. Again, this time window was chosen in order to account for the 200 ms that it takes to launch a saccade (Matin et al., 1993). We selected vowel offset, rather than vowel onset, to be the starting point of the time window, since listeners only had access to the critical duration cues on the vowel after hearing it in its entirety. Figure 3 shows participants' eye movements, calculated as the proportion of looks to the pictures of long vowel interpretation, split by Article Bias to either the long or the short vowel interpretation and Rate (fast vs. slow) in the vowel window, with the vowel analysis window shaded in grey. Figure A (Supplementary Materials) illustrates the effect of Continuum reported below.

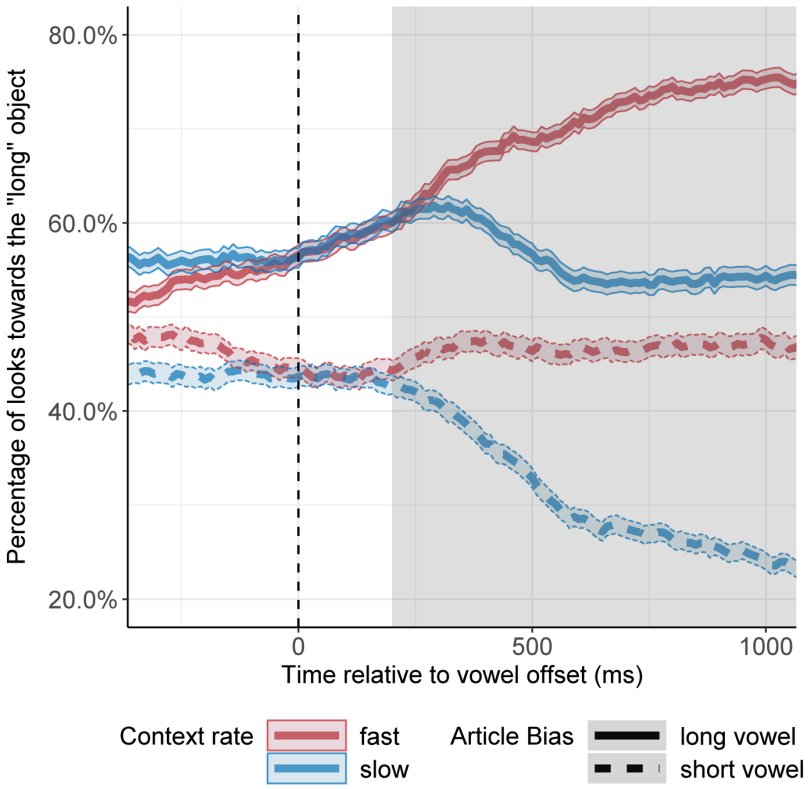


*Figure 3. Percentage of looks to the "long" object across time in the vowel window.* Time point 0 marks the offset of the ambiguous vowel. Looks towards the long item are plotted across time in two syntactic categories (trials with a 'long vowel' Article Bias contained an article that biased participants towards the object corresponding to the long interpretation: solid line; trials with a 'short vowel' Article Bias contained an article that biased participants towards the short interpretation: dashed line) when embedded in contexts of distinct rates (fast rate: red line, slow rate: blue line). The area shaded in grey indicates the window analysed in the vowel window analysis, spanning from 200 ms after vowel offset until stimulus offset. Red and blue shading indicates standard error of the mean.

A GLMM tested the binomial looks to the object corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of Article Bias (categorical predictor with two levels: long coded as +0.5; short as -0.5), Continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), Rate (categorical predictor with two levels: fast coded as +0.5; slow as -0.5), Time (z-scored around the mean of the analysis window), and all their interactions. The random effects structure contained random intercepts for Participants and Items and by-participant and by-item random slopes for all fixed factors (but not their interactions, as the model failed to converge if they were also added to the random effects structure).

The complete model outputs are summarized in Table 4 (left column "Base Model"). The model revealed significant main effects of Continuum ($p < 0.001$; Figure

A, Supplementary Materials) and Rate ($p < 0.001$). These results indicate that 1) participants were more likely to look at the picture corresponding to a long vowel at higher continuum steps, and 2) participants were more likely to look at the "long" picture in fast as opposed to slow contexts. Note that the main effect of Rate indicates that rate manipulations have an effect on vowel perception *independently* of any Article Bias. The model also revealed a significant effect of Article Bias ($p < 0.001$), indicating that participants were still more likely to look at the "long vowel" picture in the vowel window if the preceding article was congruent with a "long vowel" interpretation. It thus appears that, generally, participants did not entirely dismiss the morphosyntactic cue upon hearing the acoustic cue, nor the other way around.

On top of these main effects, the model also revealed two-way interactions between Time and Rate ($p < 0.001$), Time and Continuum ($p < 0.001$), and Time and Article Bias ($p < 0.001$). These interactions indicate that the effects of Rate, Continuum and Article Bias all grew stronger over time. The absence of an interaction between Rate and Article Bias ($p = 0.251$) further suggests that participants used knowledge- and signal-based cues independently during the vowel window (but see section "Investigating individual strategies").

There were also small significant interactions between Rate and Continuum ($p < 0.001$), indicating that rate effects were slightly stronger at higher ends of the vowel continuum; and between Continuum and Article Bias ($p = 0.018$), indicating that the effect of Article Bias was more pronounced at higher continuum steps. However, since these have relatively small effect sizes and we did not have specific predictions regarding interactions with Continuum, we do not discuss these further.

Furthermore, the model also revealed several significant three-way interactions and even a four-way interaction. Note, however, that all these interactions had very small

estimates, contributing only modestly to the observed patterns. The model revealed an interaction between Time, Rate and Continuum ($p < 0.001$), indicating that the rate effect grew slightly stronger across time at higher continuum steps; a three-way interaction between Time, Rate and Article Bias ($p = 0.006$), indicating that the rate effect grew slightly stronger across time for trials that were biasing participants towards the long interpretation; and a three-way interaction between Rate, Continuum and Article Bias ($p < 0.001$), suggesting a slightly diminished effect of Rate for trials with a long-vowel-congruent Article Bias at lower continuum steps. Finally, the model revealed a significant four-way interaction between Rate, Article Bias, Continuum and Time ($p < 0.001$); we currently lack an explanation for this, but note that the effect is very small.[2]

**Investigating individual strategies**

Taken together with previous findings by Morrill et al. (2015), Mattys et al. (2007), and Heffner et al. (2015), our results point towards a mechanism of spoken language comprehension that integrates cues from both knowledge- and signal-based levels. However, all previous studies reported averages, so it cannot be ruled out that individual participants showed strong preferences for one of the two cues. Specifically, in our study, the knowledge-based effect of Article Bias and the signal-based effect of Rate could be driven by different participants. This is especially interesting in light of our second research question: Investigating whether rate effects in online gaze patterns persist even for participants who behaviorally favor the syntactic cue will allow us to gain new insights into the robustness of phoneme-level contextual rate effects. In the

2 In order to investigate whether this effect changed as a function of experimental block, we also tested a model including an additional fixed effect of Block and interactions between ArticleBias*Block and Rate*Block. This model revealed no additional main effect of Block and no interactions with Block.

following section, we report an analysis in which we map participants' behavioral responses to their eye-tracking data.

Participants' behavior on the categorization task can be classified to fall in between two extremes, depending on which of the cues the participants weighted more strongly during the experiment. "Syntax-followers" would attribute a higher weight to the morphosyntactic information carried in the definite article, while "acoustics-followers" would weigh the acoustic information induced by the contextual speech rate more strongly. Participants' behavioral categorization responses offer insights into which of the two cues they preferred in explicit categorization, and thus by proxy into which cue they weighted more strongly. Investigating each participant's eye-tracking behavior while taking their behavioral preference into account thus yields further insights into how different participants combined the two (possibly competing) cues in an online fashion, and whether the dispreferred cue was still considered by individuals that behaviorally favored the other cue.

For further analyses, we first created an Individual Strategy variable, which captured each participant's ratio of syntax-following responses. Specifically, we calculated the proportion of each participant's "long" responses after hearing an article biasing towards a "long" response, and subtracted from this the proportion of "long" responses after hearing an article biasing towards a "short" response. This resulted in an Individual Strategy score between 0 and 1 for each participant (mean = 0.42, SD = 0.36, min = 0.03, max = 1; complete data given in Table B in Supplementary Materials). Participants that weighted the morphosyntactic cue on the article very strongly, and the contextual speech rate less so, would be expected to have an Individual Strategy score approaching 1. In contrast, participants that weighted the contextual speech rate cue more strongly, would have an Individual Strategy score around 0. Generally,

participants appeared to behaviorally favor a mixture of the two cues. This is reflected in the group mean Individual Strategy score of 0.42, as well as in the observation that no participant had an Individual Strategy score of exactly 0, and only one participant had an Individual Strategy score of 1.

For plotting purposes, we split participants into two groups based on their Individual Strategy scores. Participants with an Individual Strategy higher than the mean (0.42) were considered syntax-followers, participants with an Individual Strategy score of 0.42 or lower were considered acoustics-followers. Figure 4 shows the eye-tracking responses split by group.

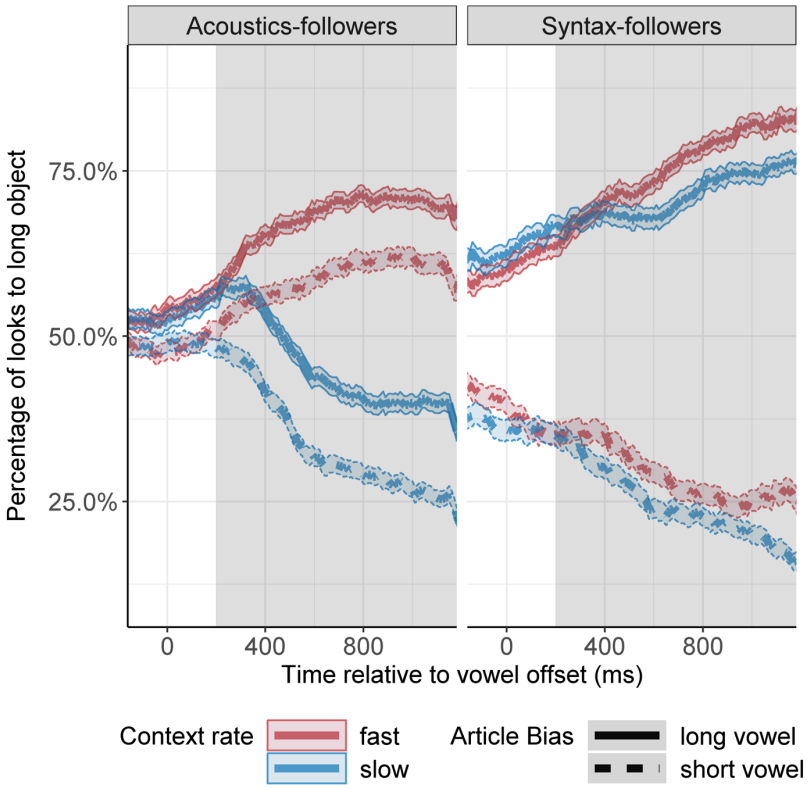


*Figure 4. Percentage of looks to the object corresponding to a long vowel interpretation across time in the vowel window, split by acoustics-following (left panel) vs. syntax-following (right panel) participants.* Time point 0 marks the offset of the manipulated vowel. Looks towards the long item are plotted across time following the onset of the noun in trials with a long Article Bias (solid line) and trials with a short Article Bias (dashed line). Ambiguous vowels were embedded in contexts at a fast (red line) or slow rate (blue line). The area shaded in grey indicates the analysis time window, ranging from 200 ms after vowel offset until the end of the stimulus. Red and blue shading indicates standard error of the mean.

We extended the GLMM which analyzed the vowel time window (described in section "Vowel window analysis") to include a main effect of Individual Strategy

(continuous predictor, centered), an interaction term between Article Bias and Individual Strategy, and an interaction term between Rate and Individual Strategy.

The complete model outputs are summarized in Table 4 (right column "Extended Model"). The model revealed the same significant effects as the simpler base model for the vowel time window, with main effects of Rate, Article Bias, and Continuum. In addition, we observed a main effect of Time ($p < 0.001$) and a small interaction between Rate and Article Bias ($p = 0.002$), which were not significant in the simpler model. These additional effects indicated that participants looked more towards the long object as time progressed, and that the overall rate effect was slightly more pronounced in trials in which the article biased listeners towards a short vowel interpretation.

Crucially, the extended model revealed an additional significant interaction between Rate and Individual Strategy ($p < 0.001$), indicating that Rate effects were stronger for participants with a lower Individual Strategy score (i.e., acoustics-followers), as well as a significant interaction between Article Bias and Individual Strategy ($p < 0.001$), indicating that the effect of Article Bias was more pronounced for participants with higher Individual Strategy scores (i.e., syntax-followers).

Taken together, these findings indicate that, while it appears to be the case that different participants employed different strategies during the experiment, participants were unlikely to rely exclusively on either of the two cues. Crucially, the effect of Individual Strategy modulated the Rate effect only to a limited extent. In fact, based on the estimates of the predictor Rate and the interaction between Rate and Individual Strategy, one learns that the model still predicts a small Rate effect for participants with an Individual Strategy score of 1 (i.e., participants that exclusively gave syntax-following responses). Specifically, recall that the estimate of Rate of 0.70 reflects the rate effect at the mean Individual Strategy score (i.e., 0.42). The estimate of the interaction between

Rate and Individual Strategy (-1.15) allows calculation of the predicted Rate effect at an extreme Individual Strategy score of 1: 0.70 + (-1.15 * (1 - 0.42)) = 0.033. Although this value is small, it still reflects a positive Rate effect, as predicted.

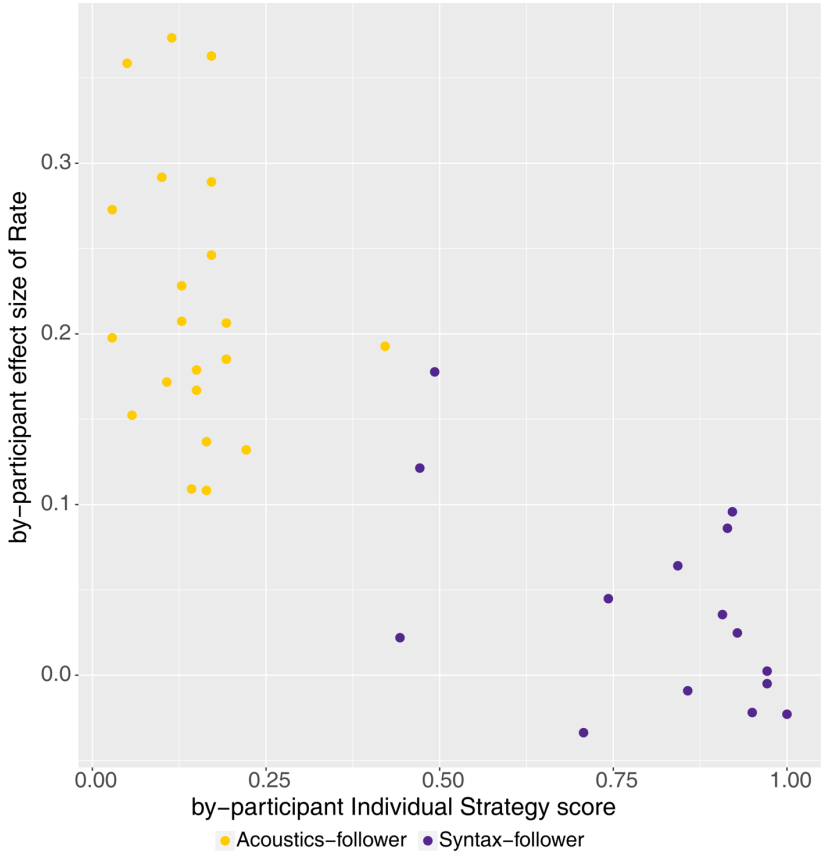


*Figure 5. Individual participants' rate effect size plotted against their Individual Strategy score.* Individual Strategy scores were calculated as the proportion of syntax-adhering responses for each participant, and rate effect sizes were calculated as each participant's difference in looks to the long object between slow and fast context sentences in the vowel window. For illustration purposes only, participants are color-coded as "syntax-followers" (purple dots; Individual Strategy score > 0.42) or "acoustics-followers" (yellow dots; Individual Strategy score <= 0.42).

In order to further illustrate this, we linked each participant's rate effect size in their looking behavior to their individual behavioral Individual Strategy score. Specifically, we calculated individual eye-tracking rate effect sizes in the vowel window by subtracting each participant's mean proportion of looks to the long object in slow contexts from their mean proportion of looks to the long object in fast contexts. The resulting measure thus captures the difference in that participant's eye fixation behavior between fast and slow contexts, and thus their individual rate effect. Figure 5 shows each individual's rate effect size plotted against their Individual Strategy score. We color-

coded participants with an Individual Strategy score equal to or higher than 0.43 as syntax-followers (purple dots) and those with a lower score as acoustics-followers (yellow dots) for illustration purposes.

## Discussion

The aim of the current study was to investigate two main questions. First, we asked whether knowledge-based, morphosyntactic cues towards gender are immediately used to generate predictions about upcoming referents. Second, we asked whether signal-based, contextual speech rate effects persist even in the presence of earlier disambiguating morphosyntactic information. We addressed these questions by experimentally inducing contextual rate normalization effects on an ambiguous vowel between short /ɑ/ and long /a:/ in Dutch minimal pairs, while at the same time providing an earlier morphosyntactic gender cue. In the following, we discuss our results in light of these two questions.

### Knowledge-based cues are rapidly taken up and used to make predictions about upcoming referents, even in the presence of uncertainty

Our analyses of the article window show that participants were more likely to look at the object corresponding to the vowel interpretation that was consistent with the morphosyntactic gender information conveyed in the definite article. These results indicate that participants rapidly use knowledge-based cues in order to make predictions about the gender of the upcoming noun.

Huettig and Janse (2016) reported results from a similar eye-tracking experiment in which participants were presented with auditory stimuli containing an article that matched only one of four possible objects on the participant's screen (e.g., *Kijk naar de*$_{\text{COMMON}}$ *afgebeelde piano*$_{\text{COMMON}}$, "Look at the displayed piano"). They found anticipatory looks to the target picture well before target onset, suggesting that

participants made predictions about the upcoming target noun. Note, however, that their experimental manipulation always included a 1:1-correspondence between the article and the following auditory target noun. The article was thus an extremely salient and reliable cue that univocally pointed to the upcoming target noun. Here, we report evidence for anticipatory language processing based on the gender of the article even though it is not necessarily a reliable cue towards the noun that followed it.

Our results, and those obtained by Huettig and Janse (2016) and others (e.g., Martin, Monahan, & Samuel, 2017; Szewczyk & Schriefers, 2013; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha et al., 2003, 2004) indicate that listeners use knowledge-based cues to predict upcoming words. As such, our results add to the recent debate about the role of predictions in language processing (e.g., Nieuwland et al., 2018). Importantly for our work, Kochari and Flecken (2019) and others (e.g., Nicenboim & Guerra, 2018) reported evidence suggesting that listeners do not necessarily predict the gender of an upcoming noun based on knowledge-based (semantic) cues. Moreover, Huettig and Guerra (2019) recently showed that the prediction of a target noun based on the gender of a preceding article could be attenuated by factors such as shorter preview time and faster speech rate of the carrier sentence. Specifically, they only observed anticipatory looks towards a target object in situations where auditory targets were preceded by a sentence presented at a slow rate. For “normal” (faster) contextual speech rates, participants appeared to only predict the upcoming material if they had ample time to preview the potential targets (long preview: 4 seconds; short preview: 1 second), or if they were specifically instructed to make predictions. Huettig and Guerra (2019) take these findings to indicate that prediction is not a necessity during language processing. In our current experiment, we do find anticipatory looks towards the target picture

based on the gender of the preceding article, both for slow *and* fast speech rates, showing that faster speech rates do not necessarily "eliminate" predictions all together. Taking the present results together with those reported in Huettig and Guerra (2019), we conclude that listeners are flexible in their use of cues and their weighting, a point which we return to below. Note that this is entirely in line with Huettig and Guerra's (2019, p. 200) conclusion that prediction is "contingent on the situation the listener finds herself in"; from a cue integration perspective, we would argue that prediction is contingent on the *reliability and weighting of the available cues*.

As Kochari and Flecken (2019) mention, an undoubtedly important objective of future research will be to investigate the content and extent of lexical predictions in more detail. For the present research, concerning the integration of different types of cues, it was important to demonstrate that the knowledge-based cue of gender marking was indeed utilized by the participants in our experimental paradigm.

**Signal-based, contextual speech rate effects persist even in the presence of preceding knowledge-based, morphosyntactic cues**

We observed more looks towards the picture corresponding to a long vowel interpretation in the vowel window for items embedded in fast context sentences. Crucially, this effect arose *in spite of* preceding morphosyntactic cues, which participants demonstrably made use of to make predictions earlier on (see previous section). This result suggests that contextual speech rate acts as a salient cue for language processing. Moreover, we did not find evidence for a differential effect size of contextual speech rate on participants' categorization decisions in the pretest (i.e., without preceding articles) vs. eye-tracking experiment (with preceding articles). This absence of an interaction between rate effects and morphosyntactic constraints corroborates earlier work (Heffner et al., 2015; Morrill et al., 2015), together

highlighting the automaticity of contextual rate effects.

In addition, the rate effect observed in the eye-tracking data arose very rapidly in time (around 200-250 ms after vowel offset; cf. Figure 3), which is about the earliest time point at which effects can be expected to emerge in eye-tracking data (Matin et al., 1993). This is in line with previous studies observing very early evidence for acoustic context effects in eye-tracking experiments (Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). Thus, our results add to a growing body of literature showing that effects of contextual speech rate are robust and arise very early during perception (e.g., Bosker, 2017; Maslowski et al., 2018).

We interpret these findings with reference to the two-stage model of acoustic context effects, introduced in Bosker et al. (2017). In this model, acoustic context effects, including rate normalization, are suggested to arise at two distinct processing stages. The first stage encompasses early and automatic perceptual normalization processes, while a second stage involves later cognitive adjustments, for instance driven by indexical speech properties. The early time point of the rate effects and the robustness across individuals together suggest that the rate effects observed here arose at the first stage of contextual processing.

For the first time, we also report individual variation in weighting the signal-based and knowledge-based cues by mapping eye-tracking onto behavioral results. We showed that even participants who had a clear behavioral preference for the knowledge-based cue for the most part still exhibited small rate effects in their eye fixations. In fact, all participants except for some individuals at the extreme end of the Individual Strategy scale showed an effect of Rate.

These findings are difficult to integrate within models of speech comprehension that posit a stronger influence of syntactic, knowledge-based cues compared to signal-

based cues such as contextual speech rate (e.g., Mattys et al., 2005). In fact, our observation that individual participants employed different behavioral strategies during the experiment challenges speech comprehension models that propose a fixed hierarchy of cues. Instead, our results suggest that cues can be weighted flexibly during the comprehension process, both on a group level and by individual participants. Language comprehension models of cue integration (Martin, 2016) offer a promising formalization that can accommodate these results.

Our data also speak to the question how the system handles uncertainty across time. We found that participants immediately used both the morphosyntactic and the acoustic cue as soon as they became available, rather than delaying looks to either of the pictures until all the information about the entire sentence was available, or disregarding one cue entirely. Instead, as suggested by cue integration frameworks, participants appeared to immediately combine the available cues in order to arrive at a robust percept.

Based on our findings, we can formulate new questions for future research. Most notably, the question arises which general factors determine the weighting and reliability of cues. Our results indicate that listeners weighted knowledge-based and signal-based cues flexibly, but what drives this cue weighting in the presence of uncertainty and across different experimental settings? It is possible that cue reliabilities are strongly modulated by exogenous, situational cues. For example, Martin (2016) suggested that non-linguistic percepts such as gaze, facial expression, or joint-action contexts might modulate the reliability of certain linguistic cues in dialogue settings. Investigating these additional factors behind cue weighting and the interplay between cue reliabilities and their underlying modulators will be an exciting objective of future research.

Our experiment investigated two specific cues: gender information conveyed by a definite article, and contextual speech rate. It is unclear how specific our findings are to precisely these two cues, and whether a different, or more nuanced, picture might emerge for other combinations of cues. As such, caution should be taken when making claims about the integration and weighting of knowledge-based and signal-based cues in general. Rather, we believe that our results are a first step towards establishing a set of cues that the system *can* draw on during language processing, and how it can combine them (Martin, 2016). Further experiments could investigate different combinations of cues in more detail in order to observe whether similar effects arise.

Our findings are particularly interesting in light of results reported by Mattys et al. (2007). They found that attenuating effects of conflicting acoustic cues on the reliability of syntactic cues were contingent on the acoustic cue being realized *before* the syntactic one, suggesting that cue reliability and weighting can be modulated by the time course and order in which different pieces of information enter the system. For our current experiment, we would argue that the realization of the acoustic cue occurred *after* the morphosyntactic cue. Although the contextual rate information was available from the beginning of the sentence, it only became meaningful upon perception of the duration of the ambiguous vowel due to the Continuum manipulation. This information always occurred after the article. In our experiment, the influence of acoustic cues on target perception were thus *not* contingent on their time course within the stimulus. We also showed that individual participants employed different strategies when weighting and integrating the acoustic and syntactic cues with each other – this is clear evidence against a strict hierarchy of cue weights. Further, our observation that small rate effects still arose for many participants with a clearly syntax-driven Individual Strategy also

demonstrates that contextual speech rate is a robust acoustic cue that is not easily "overwritten" by conflicting syntactic information.

On a related note, Reinisch, Jesse, and McQueen (2011) conducted a series of experiments in which they investigated the use of *distal* and *proximal* contextual speech rate cues. While listeners generally appeared to rely more strongly on proximal than on distal context, the results also suggested that effects of distal speech rate grew stronger with the amount of context that listeners were presented with (i.e., "longer" contexts elicited more pronounced rate effects than "shorter" contexts). Reinisch et al. (2011) interpret this as a "cumulative effect". An interesting question for future research would be to investigate in more detail which modulators cause listeners to weigh certain cues more strongly than others.

Language comprehension usually takes place in settings that are more natural and flexible than our experimental setup. Further experiments could therefore investigate the interplay of knowledge-based and signal-based cues in a more naturalistic setting, for example during dialogue. Given that we find rate effects to be robust, even in the presence of disambiguating morphosyntactic information, it would be interesting to investigate to which extent they persist in situations that more closely resemble "real life" language use, where a lot more variability exists.

Taken together, our findings indicate that listeners rapidly extract and integrate both morphosyntactic, knowledge-based cues conveyed by a definite article and signal-based, acoustic cues conveyed by contextual speech rate. Rather than processing these cues separately in a strictly hierarchical fashion, listeners appear to take all available sources of information into account and update their beliefs about the incoming speech material depending on the reliability that they assign to the available cues.

**Acknowledgements**

We would like to thank Merel Maslowski for lending her voice for the recordings.

## References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Audacity Team. (n.d.). *Audacity*.

Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2018). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-018-1626-4

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: linear mixed-effects models using S4 classes. R package (Version 0.999375-42). Retrieved from http://CRAN.Rproject.org/package0lme4

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5) [52]. Retrieved from http://www.praat.org/

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, *79*(1), 333–343. https://doi.org/10.3758/s13414-016-1206-4

Bosker, H.R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech. Behavioural evidence from rate normalization. *Language, Cognition and Neuroscience, 33*(8), 955-967.

Bosker, H. R., & Reinisch, E. (2017). Foreign Languages Sound Fast: Evidence from Implicit Rate Normalization. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01063

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176. https://doi.org/10.1016/j.jml.2016.12.002

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or Disappear. *Psychological Science*, *21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452–465. https://doi.org/10.1016/j.wocn.2009.07.006

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442. https://doi.org/10.1038/nrn3503

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics*, *43*(2), 137–146. https://doi.org/10.3758/BF03214191

Heffner, C. C., Newman, R. S., Dilley, L. C., & Idsardi, W. J. (2015). Age-Related Differences in Speech Rate Perception Do Not Necessarily Entail Age-Related

Differences in Speech Rate Use. *Journal of Speech Language and Hearing Research*, *58*(4), 1341. https://doi.org/10.1044/2015_JSLHR-H-14-0239

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kochari, A., & Flecken, M. (2019). Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. https://doi.org/10.31234/osf.io/9npue

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*, 2867–2875.

Lisker, L., & Abramson, A. S. (1967). Some Effects of Context On Voice Onset Time in English Stops. *Language and Speech*, *10*(1), 1–28. https://doi.org/10.1177/002383096701000101

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Martin, A. E. (2016). Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00120

Martin, A. E., Monahan, P. J., & Samuel, A. G. (2017). Prediction of Agreement and Phonetic Overlap Shape Sublexical Identification. *Language and Speech*, *60*(3), 356–376. https://doi.org/10.1177/0023830916650714

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2018). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000579

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. https://doi.org/10.3758/BF03206780

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 960–977. https://doi.org/10.1037/0096-1523.33.4.960

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues

computed relative to expectations. *Psychological Review*, *118*(2), 219–246. https://doi.org/10.1037/a0022325

McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. https://doi.org/10.1006/jmla.1998.2568

Mitterer, H. (2018). The singleton-geminate distinction can be rate dependent: Evidence from Maltese. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 6. https://doi.org/10.5334/labphon.66

Morrill, T., Baese-Berk, M., Heffner, C., & Dilley, L. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, *22*(5), 1451–1457. https://doi.org/10.3758/s13423-015-0820-9

Nicenboim, B., & Guerra, E. (2018). *A crack in the crystal ball: Evidence against preactivation of gender features in sentence comprehension*. Poster presented at the AMLaP Conference (Architectures and Mechanisms for Language Processing), Berlin, Germany.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*. https://doi.org/10.7554/eLife.33468

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. Journal of Experimental Psychology: Human Perception and Performance, 37(3), 978.

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116. https://doi.org/10.1016/j.wocn.2013.01.002

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314. https://doi.org/10.1016/j.jml.2012.12.002

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, *30*(5), 529–543. https://doi.org/10.1080/23273798.2014.946427

Tuinman, A., Mitterer, H., & Cutler, A. (2014). Use of Syntax in Perceptual Compensation for Phonological Reduction. *Language and Speech*, *57*(1), 68–85. https://doi.org/10.1177/0023830913479106

Van Bergen, G., & Bosker, H.R. (2018). Linguistic expectation management in online discourse processing: An investigation of Dutch inderdaad 'indeed' and eigenlijk 'actually'. *Journal of Memory and Language, 103,* 191-209.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939–950. https://doi.org/10.3758/BF03193621

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168. https://doi.org/10.1016/S0304-3940(03)00599-8

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487