Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations.

Article in press, Brain & Language

McCall E Sarrett[1,*], Bob McMurray[2], and Efthymia C Kapnoula[2,3]

[1] Interdisciplinary Graduate Program in Neuroscience, 356 Medical Research Center, University of Iowa, Iowa City, IA, 52242

[2] Department of Psychological & Brain Sciences, W311 Seashore Hall, University of Iowa, Iowa City, IA, 52242

[3] Basque Center on Cognition, Brain, & Language, Mikeletegi Pasealekua, 69, 20009 Donostia, Gipuzkoa, Spain

ABSTRACT

Understanding spoken language requires analysis of the rapidly unfolding speech signal at multiple levels: acoustic, phonological, and semantic. However, there is not yet a comprehensive picture of how these levels relate. We recorded electroencephalography (EEG) while listeners (N=31) heard sentences in which we manipulated acoustic ambiguity (e.g., a *bees/peas* continuum) and sentential expectations (e.g., *Honey is made by* bees). EEG was analyzed with a mixed effects model over time to quantify how language processing cascades proceed on a millisecond-by-millisecond basis. Our results indicate: (1) perceptual processing and memory for fine-grained acoustics is preserved in brain activity for up to 900 msec; (2) contextual analysis begins early and is graded with respect to the acoustic signal; and (3) top-down predictions influence perceptual processing in some cases, however, these predictions are available simultaneously with the veridical signal. These mechanistic insights provide a basis for a better understanding of the cortical language network.

**corresponding author* *McCall E. Sarrett*
e-mail *mccall-sarrett@uiowa.edu*

1. INTRODUCTION

Typical listeners appear to process speech effortlessly despite the significant computational challenge posed by speech perception. Speech input unfolds continuously over time, requiring rapid analysis of linguistic information at multiple levels. This analysis must be performed on a signal with substantial acoustic variability and contextual dependency. This complexity is underscored by the large number of people with communication disorders who may struggle with speech perception: Hearing loss, for example, affects up to 15% of adults, and as many as 12% of school-aged children have language-related disorders (Tomblin et al., 1997), which have been linked to speech perception and word recognition deficits (McMurray et al., 2010, Robertson et al., 2009, Werker and Tees, 1987, Vandewalle et al., 2012). This emphasizes the need to fully characterize the cortical and cognitive mechanisms by which typical listeners perceive speech.

A key question is how early perceptual processing relates to higher order (semantic, meaning-based, or broader sentential and contextual) processing. By definition, higher order processing relies on feedforward input from perception, and the ascending path of auditory input through subcortical auditory structures and cortical language areas is well-described (Hickok and Poeppel, 2007, Malmierca and Hackett, 2010, McQueen et al., 2016). However, the nature of the computations that relate acoustic input to these higher levels is not fully characterized along three crucial dimensions: 1) the time period over which information about lower levels of analysis (e.g., auditory, phonetic cues) are maintained as later steps (e.g., lexical/semantic) unfold, 2) when semantic access begins, and 3) whether there are descending (feedback) modulations of the incoming auditory signal (McQueen, Eisner, & Norris, 2016).

Addressing these questions is critical for two larger debates in language processing (and cognitive neuroscience). First, is speech perception accomplished by (1) rapidly transforming the signal into discrete units (discarding acoustic variability) (Liberman et al., 1957, Chang et al., 2010) or (2) preserving a more flexible (but perhaps noisier) gradient representation (Kapnoula et al., 2017, Port, 2007)? Second, do listeners maintain a only a veridical (bottom-up) representation of the input (Firestone and

Scholl, 2016, Norris et al., 2000, Lupyan and Clark, 2015) or is perception biased by top-down expectations (McMurray and Jongman, 2011, Getz and Toscano, 2019, Broderick et al., 2019)? The present study uses a novel electroencephalography (EEG) paradigm to address three questions concerning the dynamics of processing at different levels of speech perception.

1.1 *First, we ask how long fine-grained acoustic detail is maintained even after lexical and semantic processing has begun.* For example, Voice Onset Time (VOT) is the primary acoustic cue that distinguishes voiced and voiceless sounds, such as /b/ and /p/. It reflects the time between the opening of the articulators and the onset of vocal fold vibration. A prototypical /b/ has a short VOT of 0-15 msec, whereas a prototypical /p/ has a longer VOT of 40-60 msec, with a category boundary between 15 and 25 msec (Allen et al., 2003, Lisker and Abramson, 1964). A classic debate in speech perception is whether small differences that would still signal the same phoneme (e.g., between 50 and 60 msec, both /p/'s) are discarded or preserved. Under a categorical view, ignoring such differences may be more efficient, as they are likely noise (Liberman et al., 1957); alternatively, preserving within-category differences may allow for more flexible processing (McMurray et al., 2009).

Most recent work suggests listeners preserve such differences in a gradient manner (McMurray et al., 2002, Miller, 1997, Andruski et al., 1994). However, while this issue is resolved in psycholinguistics, there is still debate over the underlying neuroscience (Chang et al., 2010, Toscano et al., 2010). A more compelling way to ask this question is to ask what listeners might *do* with a gradient representation. Critically, for a gradient representation to be useful, it must persist long enough in time to help listeners update decisions more flexibly.

Prior studies suggest listeners maintain fine-grained detail (e.g., the degree of voicing) for 200-500 msec (McMurray et al., 2009, Gwilliams et al., 2018, Zellou and Dahan, 2019) and possibly as long as a second (Brown-Schmidt and Toscano, 2017). This is well into the period where semantic access should have begun (Gaskell and Marslen–Wilson, 1999). However, these studies generally use a "garden path" paradigm in which the auditory cue (e.g. the VOT) may be partially ambiguous, but listeners receive disambiguating lexical or sentential information later. Here, if the cue is not perceived accurately, the interpretation may be revised in light of later disambiguating acoustic information. Three issues with this paradigm prevent a clear conclusion.

First, listeners may learn (over the experiment) to maintain fine-grained detail for a longer duration than they normally would, as they repeatedly encounter situations in which their initial interpretation is wrong and is corrected by later arriving information. However, in real sentences, there may be considerably less impetus to preserve this information, as preceding sentential or other context can eliminate the early ambiguity. If representation of fine-grained acoustic detail in these contexts is preserved past the point of semantic access, this would imply that the auditory system is fundamentally organized to process the auditory input in a gradient (non-categorical) manner and to retain it over time.

Second and more importantly, in most cases, the later semantic information completely swamps effects of perceptual gradiency. As a result, it is unclear whether the initiation of semantic access may disrupt these memory representations, thus it is important to evaluate the timecourse of these processes simultaneously.

Finally, with the exception of Gwilliams et al. (2018), all of these studies measure gradiency at the lexico-semantic level (e.g., the degree of commitment to a meaning), not at the auditory level (e.g., memory for the continuous VOT value). As a result, it is unclear whether fine-grained acoustic detail is maintained in some form of perceptual memory, or simply via the relative activation of competing interpretations. Characterizing this level of analysis may require neural measures that can target auditory encoding of acoustic cues (e.g. Toscano et al., 2010).

1.2 *Second, we ask if perceptual processing continuously cascades to influence integration of semantic expectations from a sentence, or if perceptual analysis must complete before words can be integrated into the sentence.* Conventional EEG work suggests that an incoming visual stimulus (written word) or acoustic signal (spoken word) is integrated with previously-set semantic expectations in the brain about 400 msec after stimulus presentation (the canonical N400 component) (Kutas and Hillyard, 1980). This would appear to be well after perceptual processing is complete, that is, after low level perceptual information has been categorized or abstracted, when within-category detail is lost, and the resulting phoneme is passed on to the next level of processing. The typical N400 suggests a more stage-like operation in which these perceptual operations are complete and semantic processes happen later.

In contrast, recent studies using more naturalistic sentences indicates integration with meaning-based context begins at an earlier time (Broderick et al., 2018, Broderick et al., 2019, see Kutas and Federmeier, 2011 for a review, and see Dahan and Tanenhaus, 2004, for supporting evidence using the Visual World Paradigm, Tanenhaus et al., 1995). However, these studies have not simultaneously examined the timecourse of perceptual processing along with that of semantic or contextual processing. Thus, it is still unclear whether perceptual processing is complete when semantic access begins (i.e., whether the acoustic signal has been categorized), or whether perceptual-stage processing is ongoing while these contextual expectations are being used. Under the latter case, we might expect semantic processing to show sensitivity to fine-grained phonetic cues in the signal. Such a cascade has been shown for the semantics of single words (Andruski et al., 1994, McMurray et al., 2002), but not yet during the semantic processing of a sentence.

1.3 *Third, we ask if speech perception is accomplished entirely via bottom-up processing, or if top-down feedback carrying specific linguistic content plays a role*. Longstanding psycholinguistics work suggests that phoneme categorization as well as ratings of phonemic "goodness" are affected by semantic or lexical expectations (e.g. Ganong, 1980, Connine et al., 1991, Allen and Miller, 2001). However, it is unclear whether this information truly feeds back to affect perception or whether it has its effect in some form of post perceptual decision system (Norris et al., 2000, Magnuson et al., 2003, McQueen, 2003).

Cognitive neuroscience—which can target auditory encoding more directly—offers clues favoring feedback. Gow and Olson (2016) used magnetoencephalography (MEG) with a Granger Causality Analysis to demonstrate connections from higher-level language areas to lower phonological processing hubs during sentence processing. However, it is not clear whether this top-down signal reflects actual linguistic content, as opposed to less specific processes like attentional modulation (McQueen et al., 2016). Getz and Toscano (2019) overcame this using the N1 EEG component. The N1 is an early component which reflects a number of low-level auditory processes. Among several things it tracks, it shows sensitivity to the actual VOT the subject heard (Toscano et al., 2010, Sharma et al., 2000, Pereira et al., 2018, Frye et al., 2007). A more negative N1 corresponds to a short VOT, and a less negative N1 corresponds to longer VOTs. Thus, the N1 can in part reflect the content of perceptual encoding, and not just a degree of processing difficulty (as do many EEG components). This offers a way to pinpoint feedback to auditory content.

In Getz and Toscano's study, subjects saw a visual prime (MASHED) followed by an auditory target (*potatoes*). Semantic expectations (from the prime) affected the encoding of the target word's VOT (at the N1)—/b/-biased contexts showed a lower N1 than /p/-biased, even when the auditory stimulus was identical. This effect was only observed at the phonemic category boundary. This offers some evidence for feedback. However, this study used isolated auditory target words and primes with extremely high co-occurrence. Therefore, it is unclear if this effect would generalize to a more dynamic sentential context. Neither study examined the detailed timecourse of this operation to determine when it begins and how long it lasts (relative to other processes such as the maintenance of fine-grained detail). Thus, a broader investigation is necessary to establish that feedback occurs more generally, and to determine when it operates relative to the array of simultaneous processes.

1.4 *The present study.* Addressing these three questions requires us to identify the precise time windows during which processing at different levels of analysis (auditory, phonological, contextual) takes place. Conventional EEG analyses have generally focused on finding differences in limited time windows around specific component peaks. However, these types of analyses alone cannot speak to how different processing steps fit together in time. Some recent work has sought to take this more comprehensive approach in written language comprehension (see Hauk et al., 2006, which examines the timecourse of visual word recognition in EEG) and in the perception of isolated spoken words (see Ettinger et al., 2014, which examines the interaction of morphology and surprisal over time in MEG). As of yet this approach has not been applied to a level of spoken language that is arguably more dynamic and where time is perhaps more critical. Crucially, while there is a longstanding assumption that perceptual and semantic processing operate in some form of continuous cascade, there is not a comprehensive picture of how all the parts fit together over time.

Similar to these prior studies, we conducted a continuous-in-time analysis of the EEG signal during sentence processing. As in Getz and Toscano, we start from the finding that the N1 component tracks auditory encoding of acoustic cues, such as VOT. Here, an effect of VOT on the scalp voltage (lower VOTs ~ more negative N1s) is taken as a marker of perceptual encoding (particularly during the early, N1 window). We embed this paradigm in a task in which participants heard a semantically biasing sentence such as *Honey is made by—,* followed by a token from a b/p continuum (*bees/peas*), where one side of the continuum is more likely to complete the sentence. In a typical N1 paradigm, one would predict that the N1 should linearly track the VOT of the continuum with lower VOTs reflecting more /b/-like stimuli. Further, if feedback is operative, one would predict that the N1 (VOT encoding) would be lower for /b/-biased sentences, though this may be limited to more ambiguous VOTs.

To address our broader questions about the temporal aspects of language processing at different stages, we extend this paradigm to identify the timecourse of contextual/semantic processing, potential feedback signals, and the interaction of these factors with perceptual processing. These analyses provide a comprehensive picture of the timecourse of acoustic cue encoding, semantic integration, and contextual feedback during online, naturalistic sentence processing. They suggest perceptual processing of speech is fundamentally graded, and this gradiency persists well past the point of semantic integration. But they also suggest that semantic integration occurs very early and can exert feedback effects on perception in limited but robust ways.

2. MATERIALS & METHODS

2.1 *Participants.* Participants were 36 University of Iowa undergraduates, who completed the study for course credit in Elementary Psychology. Three participants were excluded for low accuracy on the behavioral task (less than 80% correct at VOT endpoints); two more were excluded due to excessive movement artifact in the EEG. The final sample included 31 participants (21 female, 9 male, 1 nonbinary/unreported; age range: 18-30 years).

2.2 *Design.* Participants listened to sentences and indicated with a gamepad whether the final, target word began with /b/ or a /p/. The target words were selected such that each set of target words (*bark/park*) contained minimal pairs representing the endpoints of a VOT continuum. VOT was varied from 0 to 60 msec, with steps every 10 msec. There were 10 total target word continua. Sentences were designed to bias listeners toward one endpoint or the other of a given continuum. For example, "Good dogs sometimes also *bark*" is /b/-biased, while "Driving in Iowa City is miserable because there's never anywhere to *park*" is /p/-biased. Sentences were normed in a separate free-response experiment to ensure that they predicted their target word more than 75% of the time, on average. See Table 1 for some example items, and Supplement S1 for full the full set of experimental sentences and items, with details on cloze probability of sentences, word frequency, average duration, and other stimulus characteristics. There were three /b/- and three /p/-biased sentences for each of the 10 continua. It was not possible to create sentences with neutral coarticulation (e.g., the /o/ in *to* or *also* of the preceding examples would partially predict a /b/ or /p/ due to coarticulation). Thus, sentences were recorded with

Table 1. Examples of experimental stimuli.

/B/-biased		*/P/-biased*	
Sentence	*Target Word*	*Sentence*	*Target Word*
She stole my doll, so I asked her to give it—	*back*	A school is to fish what to wolves is a—	*pack*
Quiet dogs sometimes also—	*bark*	Driving in Iowa City is miserable because there's never anywhere to—	*park*
She lied on the sandy—	*beach*	The state fruit of George is the—	*peach*

both sides of the continuum, and this was counterbalanced within subject (across trials). There were 10 continua, crossed with 7 steps of VOT, 2 coarticulation conditions, and 6 sentences per continuum, for a total of 840 experimental trials.

We also included 240 filler trials in which the target words were always consistent with the sentence (see Supplement S1). These sentences and target words were distinct from the experimental items. Filler words began with either a /b/ or a /p/ and were not manipulated along a VOT continuum. This was done to ensure that the majority of trials fulfilled contextual expectations (so subjects wouldn't fully ignore the sentence contexts) even as the target sentences were presented equally with each step of the continuum.

Finally, an additional 60 catch trials were used. Catch trials were designed to keep participants engaged with the semantic/contextual information of the sentences, to ensure that they were generating meaning-based expectations for the target word. In experimental and filler trials, participants need not explicitly attend to the sentences to accurately categorize the target word. In comparison, on catch trials, subjects heard sentences without an auditory target word and chose via button push which of two (visually presented) words best completed the sentence. Catch trials were randomly interleaved among experimental and filler trials, such that participants would not know from the sentence context alone which type of response (phonemic—/b/ or /p/, or lexicosemantic—*beach* or *peach*) would be needed. Catch trials consisted of one presentation of each of the 60 experimental sentences, with no repetitions. The response options for catch trials were minimal pairs (i.e. *beach* or *peach*). Participants correctly identified the semantically related target word on 99.2% of trials, indicating that they were maintaining attention on meaning-based content of the sentences and that the sentences consistently predicted their target word.

There were a total of 1140 trials (experimental, catch, and filler trials interleaved), and participants took roughly 1.5 hours to complete the task.

2.3 *Materials.* Context sentences were natural utterances of a female native English speaker, recorded in mono at a sampling rate of 44100 Hz. Each sentence was produced with both the /b/- and /p/-initial target words. For example, the female speaker produced both *A pelican has a long beak* and *A pelican has a long peak*. The final target words were then excised and both forms were used. Across trials, subjects heard both coarticulatory forms of the sentence with each step of the continuum, and the coarticulation of the sentence-final phoneme was included as a factor in our analyses. Sentences were cut in Praat from the beginning phoneme to the release of the sentence-final phoneme before the target word. Finally, we manipulated sentence duration to create 5 different versions of each sentence. This was done to ensure that any evoked EEG activity from the preceding sentences would average out when timelocking to the target word. We used the PSOLA (Pitch Synchronous OverLap and Add) function in Praat to create two slower versions (92.5% and 85% of the original duration) and two faster versions (107.5% and 115% of the original duration. Finally, the resulting sentences were later spliced onto the target words from the VOT continuum.

To create VOT continua for the target words, natural recordings from the same female speaker, recorded separately from the sentence, were used. These were artificially manipulated in Praat (Boersma, 2006). We started with a token from each endpoint of the continuum, cleaned and prepped as described. Tokens were matched for pitch and prosody as much as possible. Any aspiration that was naturally produced on the /b/ token was removed, such that the /b/ token had a true VOT of 0 msec. The /p/ token had at least 80 msec of aspiration. After choosing the endpoint tokens, we created the continua. We excised the beginning of the /b/ and replaced it with a similarly long portion of aspiration from the /p/. This was done in consecutively increasing 10 msec increments, from 0 (representing the most /b/-like sounds) to 60 msec (representing the most /p/-like sound), which resulted in seven VOT steps along the continuum.

2.4 *Procedure*. Participants gave informed consent before completing the study. Then, participants were fit with an EEG cap and escorted into the experiment booth. Participants were instructed that there would be two different types of trials: (1) They will hear a sentence and also hear a target word, and need to respond with what letter that target word started with (/b/ or /p/), or (2) they will hear a sentence that will end abruptly, and two visual target words will appear on the screen. In this case, they should choose what word best completes the sentence. Participants were made aware that these two different types of trials would occur randomly throughout the experiment.

On each trial, a black fixation cross on a white background was shown on a 26" monitor while participants heard sentences over Etymotic ER1 earphones, ending in a target word. The interstimulus interval between the end of the sentence-final phoneme to the beginning articulation of the target word was jittered around an average duration of 115 msec. After the target word finished playing, response options appeared on the screen, and the participant had 2 seconds to respond via a button press on a gamepad. When the participant responded or when maximum time (2 seconds) elapsed, the response options disappeared, and the trial advanced. There was an intertrial interval of 1.5 seconds.

2.5 *Electroencephalography.* EEGs were recorded with a 32-channel BrainVision actiCap system in an electrically-shielded sound-attenuated booth. Electrodes were placed according to the International 10-20 system, referenced online to the left mastoid, and re-referenced offline to the average of the left and right mastoids. Horizontal and vertical electrooculogram (EOG) recordings were obtained via electrodes placed approximately 1 cm from the lateral canthus of each eye and on the cheekbone directly below the center of the left eye. Recordings were collected via BrainVision active electrodes and the signal was amplified via BrainVision actiCHamp system. Impedance was less than 5 kΩ at all sites at the beginning of the recording session. Continuous data was resampled to 500 Hz and band pass filtered from 0.1 Hz to 30 Hz with an 8 dB / octave rolloff. Eye blinks were removed using ocular correction Independent Component Analysis (ICA), and trials with artifacts that exceeded a 100 µV change in a 100 msec window were excluded (approximately 5% of trials). Event-related potentials are timelocked to the onset of the sentence-final target word, data are baselined for each trial using a 300 msec window preceding the target word onset, and are then averaged across frontocentral electrodes (Fz, F3, F4, Cz, C3, C4). Full scalp topographies over time available in Supplement S7.

RESULTS

3.1 Semantic bias of the preceding sentence shifts listeners' categorization of the target word.

Figure 1 (left panel) shows the response data as a function of sentence context. For short VOTs, listeners were more likely to choose /b/, with a steep transition to /p/ around 30 msec. We used a logistic mixed effects model to determine whether Bias and VOT significantly influenced phoneme categorization responses. Fixed effects included Bias and Coarticulation (which were contrast coded: -1 = /b/, 1 = /p/) and VOT was a continuous predictor (scaled and centered from -1 to 1). Random effects structures were chosen through forward model selection (Matuschek et al., 2017) and included a random intercept of Subject, as well as random slopes of Bias and VOT on subject; random intercept of Item and a random slope of Bias on Item. Full details are in Supplement S3.

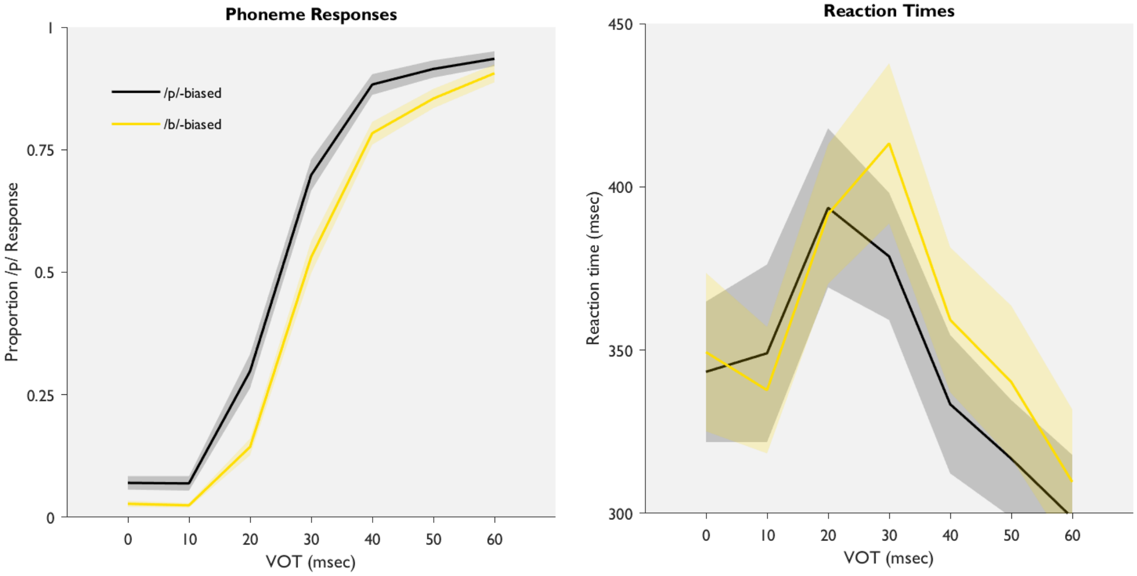

Figure 1. Behavioral response data. The left panel shows the effect of sentence context on phoneme categorization (N = 31, p = .0005). The shaded region is standard error of the mean. The right panel shows the effect of sentence context on reaction times (N = 31, p = .01).

We found a significant main effect of VOT on phoneme responses (β = 5.36, SE = 0.30, z = 18.15, *p* < .0001) with more /b/ responses for shorter steps. Sentence bias significantly shifted the categorization curve in the predicted direction (β = 0.47, SE = 0.13, z = 3.48, *p* = .0005): When listeners heard a sentence which biased them to expect a /p/-initial word, they were significantly more likely to respond /p/. This replicates previous psycholinguistic work (Miller et al., 1984) and validates the stimuli and task. Coarticulation in the final phoneme of the sentence context also played the predicted role, with more /p/ responses after a /p/-coarticulated phoneme (β = 0.37, SE = 0.02, z = 17.07, *p* < .0001; see Supplement S2 for a plot of the Coarticulation effect on categorization responses).

Reaction time (RT) data is shown in Figure 1 (right panel), and shows the expected peak near the category boundary. RT was log scaled and analyzed with a linear mixed model with fixed effects of Bias (contrast coded) and VOT and the quadratic effect of VOT (as continuous predictors). The quadratic term (VOT^2) was included since RTs typically peak during ambiguous VOTs and decrease at either category side (Pisoni and Tash, 1974). Details of this model are in Supplement S3. The model also included interaction terms of Bias x VOT and Bias x VOT^2. Random effects were selected by forward model selection and included random intercepts of subject and item, and random slopes of Bias, VOT, and VOT^2 by subject and by item.

We found a significant effect of VOT (β = -.029, SE = 0.03, t(10.19) = -1.37, *p* = .007), with shorter VOTs resulting in longer RTs, and longer VOTs resulting in shorter RTs. There was also a significant effect of VOT^2, where VOTs at the category boundary resulted in longer RTs than those at the endpoints (β = -.039, SE = 0.005, t(22.74) = -7.36, *p* < .001). Finally, we found a significant effect of the interaction Bias x VOT^2 (β = .0068, SE = 0.003, t(23370) = 2.57, *p* = .01). This reflects the fact that the peak RT shifted along with the category boundary, depending on the sentence bias.

3.2 Characterizing the comprehensive timecourse of speech processing

Figure 2 shows average voltage at frontocentral electrodes over time as a function of VOT. The gray inset zooms in on the time window from 125 to 225 msec, at the canonical N1. Longer, /p/-like VOTs result in a less negative N1, whereas shorter, /b/-like VOTs yield a more negative N1, replicating prior work. However, visual inspection of the full timecourse, also suggests a similar linear effect just after the canonical P2 component, and again late in the epoch around 700-900 msec (indicated with arrows on Figure 2). This underscores a need to expand the analyses to the full timecourse of processing, as not all of the effects of interest occurred directly at a component peak. Moreover, voltage at any moment is likely a product of multiple factors (and their interactions). Thus, characterizing processing may require an analysis that considers multiple simultaneous predictors over time.

Therefore, we developed a statistical approach to ask when each experimental manipulation (e.g., VOT, sentential context) significantly affected neural processing (the EEG waveform) over the full timecourse of processing. We ran a linear mixed-effects (LME) model every 2 milliseconds from 100 msec pre-target word onset to 1000 msec post-onset to predict the measured voltage at frontocentral electrodes from all of the experimental factors (adjusting for the large number of comparisons). From this, we determined the time regions over which a given factor (e.g., VOT) significantly impacted voltage. This approach allows us to characterize the timecourse of processing when both lower level and higher level information is available (see also Broderick et al., 2018, Broderick et al., 2019).

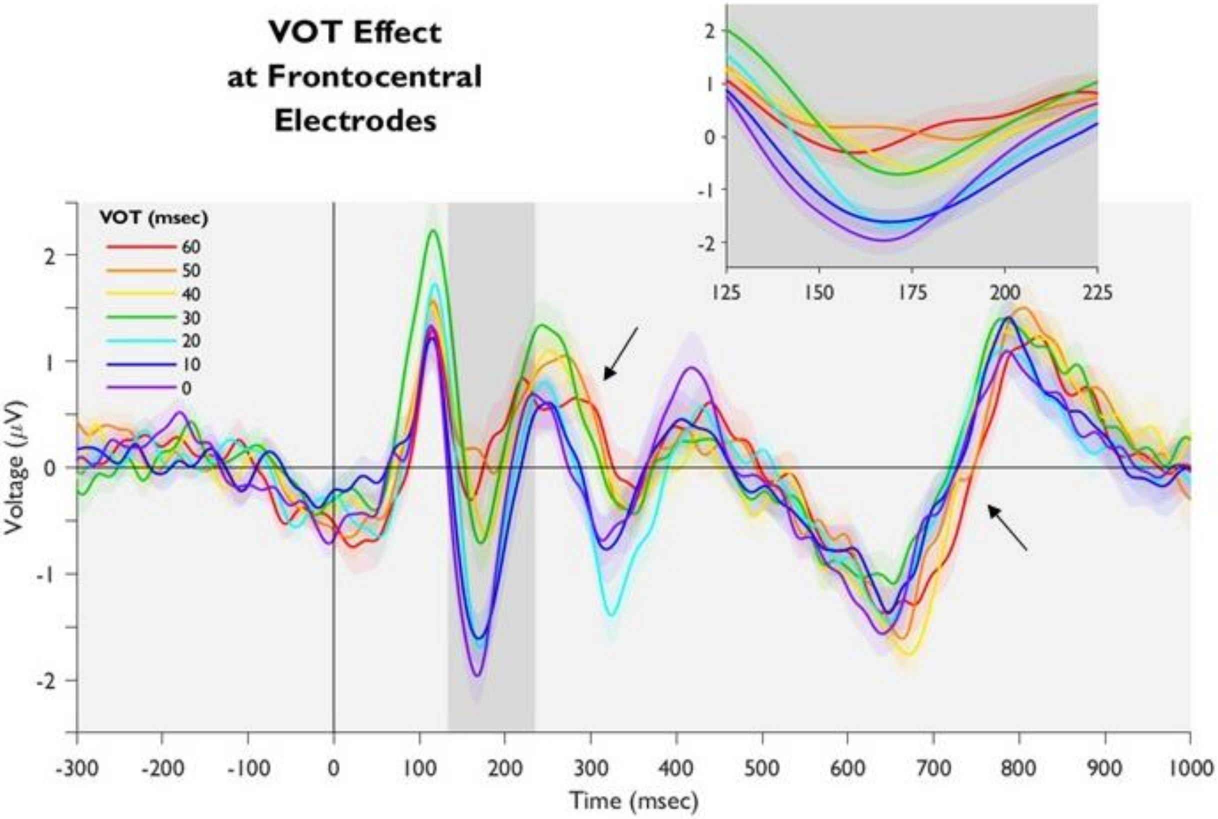

Figure 2. Averaged EEG waveform as a function of VOT, timelocked to the presentation of the target word (N = 31).

To answer our three primary questions, we examined subsets of the manipulated factors. First, to assess the duration over which fine-grained acoustic information is maintained, we examined the temporal extent of the effect of VOT, reflecting the veridical acoustic signal. Additionally, we included a quadratic effect of VOT (centered-VOT^2), reflecting an alternative hypothesis in which the voltage on the scalp does not reflect the raw VOT, but rather it reflects whether the VOT clearly belongs to a specific phonemic category or not (e.g., higher voltage at the boundary, analogous to the effect shown in reaction time literature).

Second, to identify time windows in which expectations from the sentence impacted processing, we treated VOT in terms of distance from the expected category. For example, in a /b/-biased sentence, a VOT of 0 msec (/b/-like) was perfectly expected, while a VOT of 60 msec (/p/-like) would be unexpected; conversely, for a /p/-biased sentence, a VOT of 0 msec (/b/-like) was not expected, while 60 msec (/p/-like) would be highly expected. This was operationalized as a Bias × VOT interaction. By estimating the onset of this effect, we could determine when the incoming signal is compared to sentence-based expectations. This interaction is analogous to typical N400 analyses (Kutas and Hillyard, 1980), as it reflects the degree of expectation or violation. If fine-grained representations of VOT are preserved even after the onset of semantic processing (Question 1), the linear effect of VOT should be preserved even after the Bias × VOT interaction becomes significant.

Finally, to identify feedback from sentence processing to the level of VOT encoding, we assessed a main effect of Bias. That is, if context-based expectations influence how listeners encode VOT from the earliest moments of target word processing, a /b/-biased sentence should lead to a lower voltage overall in the EEG (shorter, /b/-like VOTs = more negative N1s). However, Getz and Toscano (2019) observed this effect only at ambiguous VOTs. In our model, this should appear as an interaction of the quadratic effect of VOT and sentence bias (Bias × VOT^2 interaction). Thus, feedback would be indicated by either a main effect of Bias or a Bias × VOT^2 interaction. Moreover, if we find a time window in which VOT alone affects the signal, this suggests a point where only bottom-up acoustic information is processed, independently of top-down feedback.

To summarize, the fixed effects included the linear effect of VOT (of the target word, centered and scaled from -1 to 1), the quadratic effect of VOT (VOT^2, calculated from the centered VOT term and then re-centered and scaled from -1 to 1), Coarticulation (of the sentence-final phoneme, -1 = /b/, 1 = /p/), and semantic Bias (of the preceding sentence; -1 = /b/, 1 = /p/). We also included interaction terms for Bias × VOT and Bias × VOT^2.

Potential random effects included Subject and Item. As all of the fixed effects were within-subject and within-item, they were potential random slopes. There is some current debate around the optimal approach to selecting random effects structure (Barr et al., 2013, Matuschek et al., 2017, Seedorff et al., submitted). While Barr et al. (2013) recommends a maximal approach with fully random slopes, this approach has been shown to be anti-conservative when the data do not warrant this (Seedorff et al., submitted, Matuschek et al., 2017). The maximal approach is further complicated by the complexity of the models used here and the fact that they needed to be run hundreds of times—non-convergence would be a real problem. However, Seedorff et al. (submitted) recently introduced a model-space approach in which all possible random effects structures are tested, and the model with the lowest Akaike's Information Criterion (AIC) is kept. They showed that this holds Type I error at .05 and maximizes power, as it chooses models only as complex as necessary for the data.

Thus, random effects structure was determined applying this procedure to representative timepoints. The model with the lowest AIC was recorded at each timepoint. We then examined the distributions of models across time to select a single random effects structure that could be used for the entire time course (i.e. the model that had the lowest AIC at the highest number of timepoints). The final, random effects structure included a random intercept of Subject, random intercept of Item, and random slope of Bias on Item.

Because we were conducting multiple comparisons across time, it was important to control for family-wise error. Significance tests in a timeseries are highly autocorrelated, and thus do not represent truly independent tests. We corrected our significance tests using a true family-wise error correction developed as part of the Bootstrapped Differences of Timeseries (BDOTS) approach to timeseries analysis. This family-wise error correction that computes autocorrelation among the statistics and identifies an adjusted alpha value for each factor. This approach is less conservative than a traditional Bonferonni correction. It also offers a true family-wise error correction, unlike FDR which admits some probability of false discovery (see Oleson et al., 2017, for a derivation and Monte Carlo simulations, and Seedorff et al., 2018, for application to Visual World Paradigm data). Corrected alphas for each factor are reported in the Supplement S3. The LME model was fit in R using the lme4 package. Further model details are also available in the Supplement S3. We summarize the major findings below.

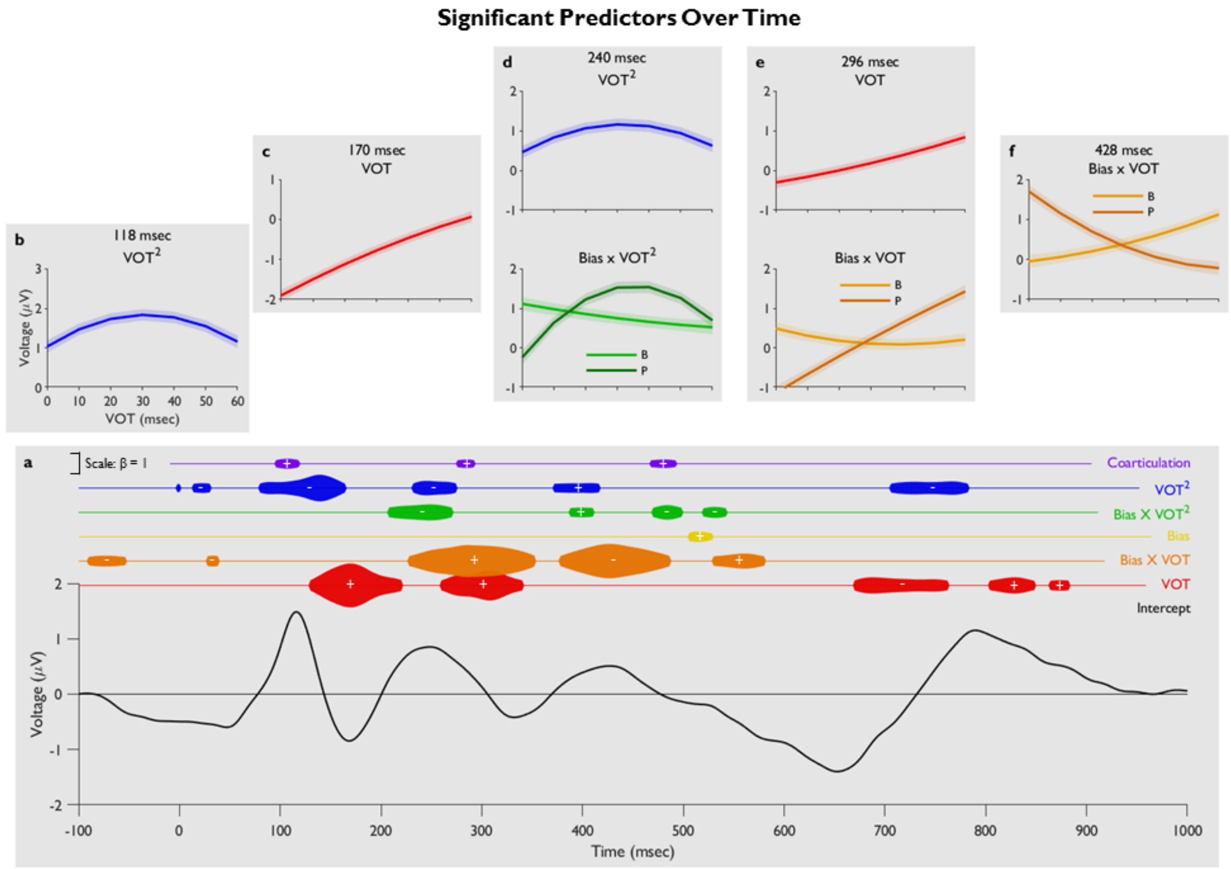

Figure 3. (A) The full model output over time. The intercept (estimated voltage from the model output) is shown on the bottom line graph with beta coefficients of significant predictors shown above; the width of the balloon corresponds to the strength of the effect and the positive/negative symbol corresponds to the direction of the effect. (B-F) Effects taken from peak effect timepoints show change in predicted voltage as a function of different predictors, calculated from a parametric bootstrap on the estimated model that time. This bootstrap estimated the predicted voltage for a "new" subject. Individual effects (like VOT) were set to the original values used in the model; for effects not shown, corresponding IVs were set to 0. Standard error of the model's predicted value is shown by the shaded region. In the calculations for B-F, the terms for VOT and VOT^2 move together, as they reflect different polynomial transformations of the same variable. VOT^2 shows the effect of phonemic ambiguity (A and D). VOT is acoustic cue encoding (C and E). Bias × VOT^2 shows the differential effect of predictions from the sentence Bias depending on whether the incoming VOT is near category boundary (ambiguous) or not (D). Bias × VOT is the integration of semantic/contextual information with the incoming spoken word (E and F).

3.3 Early perceptual processing is sensitive exclusively to the bottom-up signal, and acoustic representations endure for a long time.

Figure 3 shows the full output of the LME model, with insets showing main effects and interactions at peak effect timepoints (see Supplement S4 for a table summarizing complete results). We found an early linear effect of VOT in the time associated with the canonical N1 shown in Figure 2. This linear effect was significant from 130-220 msec (see Figure 3C, showing model-predicted voltage as a function of VOT at that time point). However, the linear effect of VOT (in red) also re-appears at multiple later timepoints with a second later window from 260-340 msec, and even later effects past 700 msec, as can be seen in Figure 2 and 3A, and in Supplement Figure S4. This suggests that fine-grained differences in the acoustic cue is maintained in cortical activity for substantially longer than they exist in the auditory signal.

The quadratic effect of VOT (VOT^2) was also significant (Figure 3A, in blue; see Supplement S6 for EEG waveform), with an extremely early effect from 80-164 msec (during the canonical P50). Like with VOT, this effect re-appears at later timepoints. Our early effect replicates Gwilliams et al. (2018); however, the later effects have not been observed previously. This suggests that multiple concurrent representations (both categorized and veridical) of the acoustic signal are maintained during processing.

3.4 Semantic processing overlaps with early perceptual analysis and is graded with respect to the fine-grained acoustic signal.

To visualize the effect of semantic expectations, we recoded VOT in terms of its distance of the expected phonemic category (i.e. /b/ or /p/). Figure 4 shows the EEG as a function of this relative VOT measure. It suggests a time just after the P2 (the dark gray inset) in which expectations affect scalp voltage, with less expected stimuli leading to a lower voltage. In our statistical model, the Bias × VOT interaction (orange in Figure 3A) was significant from 228-352 msec, corresponding with the inset in Figure 4, which is substantially earlier than typical N400 window for semantic integration, but consistent with recent work (Broderick et al., 2018). Furthermore, the effect of context integration was not restricted to this time window, but extended through about 600 msec.

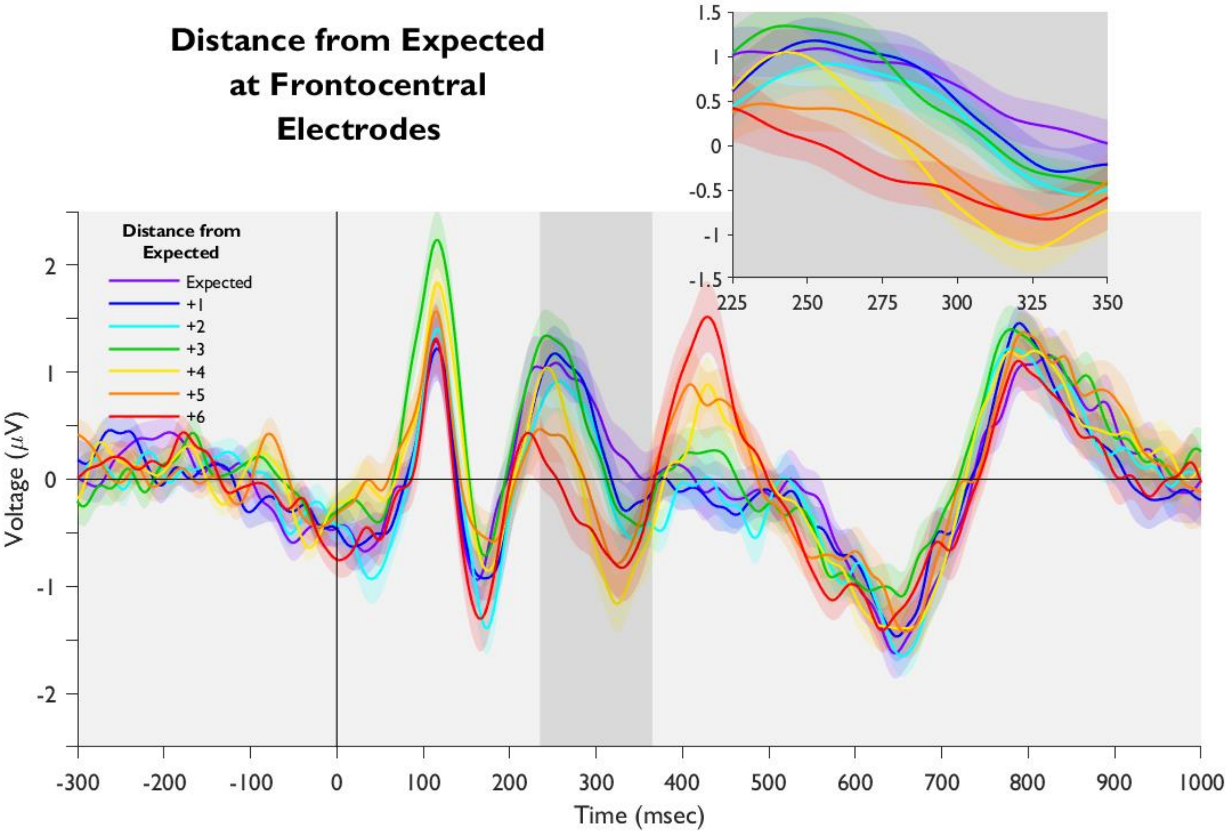

Figure 4. Averaged EEG waveform as a function of distance from expectation, timelocked to the presentation of the target word (N = 31).

Importantly, this contextual integration effect was gradient, relative to VOT. This suggests interlevel interactions and continuous cascades in processing: processing at the semantic or sentential levels interacts directly with perceptual analysis of the acoustic signal. This interpretation is further supported by the fact that the effect of cue encoding (the VOT main effect) and the effect of integration (the Bias × VOT interaction) were significant during overlapping time windows (260-340 msec and 228-352 msec, respectively).

Secondly, the direction of the interaction switches throughout processing (Figure 3E vs 3F). At early points (Figure 3E), this is consistent with the direction predicted by the conventional N400; it then reverses (Figure 3F) for about 100 msec and returns to the predicted direction (the third region, 530-580 msec). This temporary switch in direction may reflect the integration of different types of processing, such as novelty P3 effect (see Friedman et al., 2001 for a review), in which a more surprising (or less expected) stimulus yields a higher amplitude positive peak. Or, it may reflect activity originating in a different cortical area; this is considered further in the discussion.

3.5 Evidence for feedback.

There was little overall evidence for a main effect of Bias. This suggests that expectations from sentence context may not generally bias VOT encoding. However, we did find a significant Bias × VOT^2 interaction from 208-270 msec (green in 3A) and at multiple later points in the timecourse of processing. At 240 msec (Figure 3D, bottom), when listeners expect a /p/, VOTs are encoded less negatively (more /p/-like) but only at the ambiguous VOTs (in the middle of the continuum). Similarly, when listeners expect a /b/, VOTs are shifted more negatively (more /b/-like), but again only in the middle of the continuum. This effect does not hold for endpoint VOTs clearly belonging to a phoneme category, suggesting that those VOTs at category boundary may be more susceptible to some types of top-down influences.

3. DISCUSSION

This study aimed to characterize how the neurophysiological correlates of distinct components of language processing (perceptual encoding, semantic expectations, feedback) relate during real-time dynamic sentence processing. We asked (1) how long fine-grained acoustic detail is maintained in the neural signal, (2) how semantic processing relates to perceptual processing, and (3) how and when semantic feedback affects acoustic encoding. Our statistical approach used an LME model over time to capture cascading levels of processing: from acoustic encoding to graded semantic/contextual processing.

4.1 Limitations.

The study has two limitations worth discussing. First, it was necessary to have multiple repetitions of each sentence, in order to have sufficient power to detect an effect of VOT and maintain a balanced design. However, the repetition of sentences may have caused participants to engage with the sentence contexts in an unnatural way, due to task demands, or may have created stronger expectations than what would be observed in free-running, conversational speech processing. Nevertheless, this design is similar to others which have explored issues of subphonemic and sentential processing (Gow Jr and Olson, 2016, Getz and Toscano, 2019), and yet still overcomes some of the limitations of those designs. Moreover, with over 60 sentences it would have been difficult for participants to store much in memory.

Second, we limited our LME analyses to an average of frontocentral electrodes in order to capture both auditory and semantic/contextual information in the EEG signal. Thus, it could be possible that our effects are unique to this subset of electrodes. However, Supplement S5 shows a replication of our LME analysis at centroparietal electrodes to test the robustness of our conclusions. This shows a highly similar pattern of results, suggesting these effects are not an artifact of the recording sites that were chosen.

4.2 Listeners retain fine-grained representations of perceptual cues.

We found a linear effect of VOT on voltage at frontocentral electrodes at times that extend to roughly 350 msec post stimulus onset. This effect also re-appeared later in the epoch from approximately 650 msec to 900 msec, switching directions briefly from ~650-750 msec. Even this initial period is well beyond the typical N1, and in total both effects represent a substantial extension to earlier studies (Getz and Toscano, 2019, Toscano et al., 2010). It suggests the brain retains gradient differences in VOT for a considerable amount of time, long after semantic processing has begun (indicated by our Bias × VOT effect beginning at 228 msec). This long-term storage of fine-grained detail may be crucial for allowing listeners to revise earlier perceptual decisions on the basis of subsequent context. In this vein, prior psycholinguistic work suggests that listeners retain and use such information as late as one second *after* the target word (Szostak and Pitt, 2013, Connine et al., 1991, Brown-Schmidt and Toscano, 2017). Critically, we extend this by showing that listeners retain subphonemic information even when no further disambiguating linguistic information is coming. That is, even though the maintenance of VOT will not be needed to help resolve any prior ambiguity, listeners appear to spontaneously retain it anyways. This offers some of the strongest evidence to date that long-term retention of fine-grained detail is a typical mode of language processing.

One possible mechanistic basis of this effect is are actively maintaining subphonemic acoustic information (e.g., the continuous VOT) to support this kind of flexible updating. This may reflect something akin to an echoic store that could be crucial in cases when lexical analysis completely fails (e.g., in challenging signals) and the signal essentially must be reparsed. Alternatively, it could take the form of a set of continuous cue values (e.g,. VOTs) rather than a raw acoustic store. Critically, this maintenance of the VOT occurred concurrently with semantic processing and integration, and both processes overlapped at 260-340 msec. This suggests that semantic integration does not cause perceptual analysis to cease.

Another possible interpretation is that these effects do not reflect acoustic-level information from brain activity in the later time periods. The early effect of VOT around 100-200 msec almost certainly reflects perceptual-level processing, however, it is possible that this information might be rapidly abstracted to a weighted phonemic, or even lexico-semantic, representation. That is, it could be that, for example, the acoustic-level VOT of 60 msec may instead be maintained at the phonemic level as "14% /b/-like, 86% /p/-like" (or at the lexical level as "14% *beach*-like, 86% *peach*-like). Or, similarly, this later effect could be reflective of some sort of weighted motor planning or response-based representation.

Parsimony argues more for the former than the latter explanation. Particularly in the ~100-350 msec time window, it is more likely the representation of VOT itself (either as an auditory or cue-value based), than something else. In this window, the effect is strong and consistent in its polarity, At the later 650-900 msec time window, this argument may hold true. The brief switch in the direction of the VOT effect from ~650-750 msec could also indicate that the underlying neural substrate has changed, and perhaps by extension that the nature of the graded representation has changed into one of these more abstracted forms such as weighted phonemic or lexical representation.

This alternative cannot be ruled out from this data alone. However, we note that even if this late effect reflects something higher order, the VOT would be recoverable from this weighted higher-level representation. That is, if the listener knew the sound was 60% *peach-like* they could determine that VOT was likely around 25 msec. Therefore, fine-grained acoustic detail is still maintained and recoverable from net state of activity, even if the representation is maintained in a more abstracted form.

4.3 Some VOTs are more difficult than others.

In addition to the linear trend of VOT, we also found an extremely early quadratic VOT effect from 80-164 msec, and this reappeared later throughout the timecourse of processing. This effect reflects differential processing for ambiguous VOTs (in the middle of the continuum) than clearer, endpoint VOTs. It can be interpreted in at least two ways. First, this VOT^2 effect could reflect rapid, near immediate phonological categorization, suggesting a form of categorical perception (Liberman et al., 1957, Chang et al., 2010). This interpretation would appear to conflict with our claims of a robust gradient effect. However, if the VOT^2 effect does reflect a category-based response, we note that these categorized representations must be maintained in parallel to the cue level representations (Pisoni and Tash, 1974), because the linear representation of VOT is maintained concurrently (and in fact, even longer). Thus, this is not consistent with classic accounts of categorical perception.

Alternatively, the VOT^2 effect may not reflect categorization at all. Instead, it may reflect the difficulty in encoding a given VOT: non-canonical VOTs near the category boundary are harder to identify from the signal due to their relative rarity in a spoken language. This explains the longer lasting and later linear effect of VOT: after this initial encoding difficulty is overcome, the linear representation can be maintained. This would support a statistical learning account (Maye et al., 2008, Maye et al., 2002), in which listeners have less experience with category-boundary VOTs in their native language. This is also supported by neural modeling suggesting that auditory feature maps (e.g., of cues like VOT) are sensitive to the relative frequency of individual cue values, devoting more "neural representational space" to more canonical values (Salminen et al., 2009, Herrmann et al., 1995, Gauthier et al., 2007).

4.4 Contextual Integration and the N400.

We found that the electrophysiological signal in the conventional N400 time window is continuous with respect to the fine-grained acoustics, suggesting that fine-grained variation in acoustic cues (which should have been discarded during perceptual processing) cascades to affect how words are integrated into a sentence. This has not been observed in prior N400 studies (which typically do not manipulate the acoustic or perceptual form of the stimulus). These findings suggest that not only is the cue level of representation continuous with respect to the input, but this gradiency is preserved throughout the process of integrating words into sentences. That is, the degree of difficulty with which semantic alternatives are integrated with expectations from the surrounding sentential context is graded relative to the fine-grained acoustics: the further an acoustic cue is from the expected acoustic cue, the larger N400-like effect we observe.

Contextual integration began as early as 228 msec into the target word, earlier than the canonical N400, but similar to Broderick et al. (2018). This is during the time in which raw VOT is maintained in the signal, which suggests that multiple levels of linguistic analysis are taking place simultaneously. Semantic integration does not wait for perceptual processing to finish.

4.5 Context Driven Feedback and Predictive Coding.

Finally, we show some evidence for top-down feedback on the parsing of the acoustic signal. While classic debates focus on the broader question of whether feedback is present or not (e.g., Norris et al., 2000), our results suggest a more complex story. Our data show little overall effect of sentential bias on overall voltage. However, we did observe a Bias × VOT^2 interaction that is consistent with the predicted effect (and similar to Getz and Toscano, 2019). This suggests that the effect of semantic expectations on encoding of VOT may be limited to ambiguous VOTs. Perhaps it is recruited due to the difficulty in encoding these VOTs indicated by the main effect of VOT^2. This limited effect is also consistent with Allen and Miller (2001) who showed that goodness ratings can be influenced by lexical status only at the boundary, whereas speaking rate (a bottom up factor) affects them at all VOTs.

The feedback effect observed here was somewhat late in processing. This timecourse is not entirely consistent with Getz and Toscano (2019), who found an effect of bias at roughly 75-125 msec, at the peak of the canonical N1. In contrast, we did not observe this effect until around 200 msec. It may take longer for expectations set in a dynamic sentence context to build and influence perception than it might in the highly associated word pairs used by Getz and Toscano (2019). Importantly, our feedback effect still occurs before the onset of semantic integration of the target word (as shown by the Bias × VOT effect). This suggests that at ambiguous VOT steps, sentential bias plays a significant role in parsing the acoustic signal even before semantic integration of the unfolding word has begun. This is strong evidence that we are detecting a true feedback signal in the language system, and not some other type of attentional process.

At the same time, we also observed an early time window (130-220 msec) which appears to be exclusive to bottom-up perceptual processing. This type of delayed interaction between perceptual processing and top-down feedback is not fully captured by any of the current theories of speech perception. It may be consistent with a hybrid of an interactive-activation (e.g., McClelland and Elman, 1986) and a predictive-coding model (Rao and Ballard, 1999, McMurray and Jongman, 2011, Blank and Davis, 2016). In predictive coding accounts, listeners must maintain both a raw encoding of the signal (e.g., VOT) and a representation of the expected signal (e.g., the VOT that is predicted by the context). This dual representation is necessary as speech perception is not based solely on either the bottom-up signal or the expected value, but rather is based on a comparison between the two (e.g., is this VOT higher or lower than would be expected) (McMurray and Jongman, 2011). In our experiments, the earliest time window (130-220 msec), which was only sensitive to VOT, may reflect the veridical signal, while the later VOT^2 × Bias region (208-270 msec) reflects expectations. Maintaining these concurrent representations may help listeners avoid over-relying on the biased representation, which, in rare cases, may turn out to be false (Norris et al., 2000). Further experiments will be required to clarify exactly how bottom-up and top-down feedback interact in the timecourse of speech processing and whether these computations are carried out simultaneously in the same brain region or different brain regions in parallel.

As a whole, this suggests a model of speech processing in which there are no clearly delineated states and multiple processes—signal based, feedback based, and semantic integration—occur simultaneously in overlapping waves. At every level, the system is highly gradient with respect to the incoming signal, while also driven by expectations.

ACKNOWLEDGEMENTS

This work was supported by NIH grant DC008089 awarded to BM. This work was partially supported by the Basque Government through the BERC 2018-2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490, as well as by a postdoctoral grant from the Spanish Ministry of Economy and Competitiveness (MINECO; reference FJCI-2016-28019), awarded to EK. We would like to thank Thomas Farmer for helpful discussions and advice during the development of the experiment.

COMPETING INTERESTS

The authors state that we have no competing interests, financial or otherwise.

References.

ALLEN, J. S. & MILLER, J. L. 2001. Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics,* 63**,** 798-810.

ALLEN, J. S., MILLER, J. L. & DESTENO, D. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America,* 113**,** 544-552.

ANDRUSKI, J. E., BLUMSTEIN, S. E. & BURTON, M. 1994. The effect of subphonetic differences on lexical access. *Cognition,* 52**,** 163-187.

BARR, D. J., LEVY, R., SCHEEPERS, C. & TILY, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language,* 68**,** 255-278.

BLANK, H. & DAVIS, M. H. 2016. Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology,* 14**,** e1002577.

BOERSMA, P. 2006. Praat: doing phonetics by computer. http://www.praat.org/.

BRODERICK, M. P., ANDERSON, A. J., DI LIBERTO, G. M., CROSSE, M. J. & LALOR, E. C. 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology,* 28**,** 803-809. e3.

BRODERICK, M. P., ANDERSON, A. J. & LALOR, E. C. 2019. Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *Journal of Neuroscience***,** 0584-19.

BROWN-SCHMIDT, S. & TOSCANO, J. C. 2017. Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience,* 32**,** 1211-1228.

CHANG, E. F., RIEGER, J. W., JOHNSON, K., BERGER, M. S., BARBARO, N. M. & KNIGHT, R. T. 2010. Categorical speech representation in human superior temporal gyrus. *Nature neuroscience,* 13**,** 1428.

CONNINE, C. M., BLASKO, D. G. & HALL, M. 1991. Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *Journal of Memory and Language,* 30**,** 234-250.

DAHAN, D. & TANENHAUS, M. K. 2004. Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30**,** 498.

ETTINGER, A., LINZEN, T. & MARANTZ, A. 2014. The role of morphology in phoneme prediction: Evidence from MEG. *Brain and language,* 129**,** 14-23.

FIRESTONE, C. & SCHOLL, B. J. 2016. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and brain sciences,* 39.

FRIEDMAN, D., CYCOWICZ, Y. M. & GAETA, H. 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews,* 25**,** 355-373.

FRYE, R. E., FISHER, J. M., COTY, A., ZARELLA, M., LIEDERMAN, J. & HALGREN, E. 2007. Linear coding of voice onset time. *Journal of Cognitive Neuroscience,* 19**,** 1476-1487.

GANONG, W. F. 1980. Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance,* 6**,** 110.

GASKELL, M. G. & MARSLEN–WILSON, W. D. 1999. Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science,* 23**,** 439-462.

GAUTHIER, B., SHI, R. & XU, Y. 2007. Learning phonetic categories by tracking movements. *Cognition,* 103**,** 80-106.

GETZ, L. & TOSCANO, J. 2019. Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological science*.

GOW JR, D. W. & OLSON, B. B. 2016. Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, cognition and neuroscience,* 31**,** 841-855.

GWILLIAMS, L., LINZEN, T., POEPPEL, D. & MARANTZ, A. 2018. In Spoken Word Recognition, the Future Predicts the Past. *Journal of Neuroscience,* 38**,** 7585-7599.

HAUK, O., DAVIS, M. H., FORD, M., PULVERMÜLLER, F. & MARSLEN-WILSON, W. D. 2006. The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage,* 30**,** 1383-1400.

HERRMANN, M., BAUER, H.-U. & DER, R. 1995. The "perceptual magnet" effect: A model based on self-organizing feature maps. *Neural computation and psychology.* Springer.

HICKOK, G. & POEPPEL, D. 2007. The cortical organization of speech processing. *Nature reviews neuroscience,* 8**,** 393.

KAPNOULA, E. C., WINN, M. B., KONG, E. J., EDWARDS, J. & MCMURRAY, B. 2017. Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance,* 43**,** 1594.

KUTAS, M. & FEDERMEIER, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology,* 62**,** 621-647.

KUTAS, M. & HILLYARD, S. A. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science,* 207**,** 203-205.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S. & GRIFFITH, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology,* 54**,** 358.

LISKER, L. & ABRAMSON, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word,* 20**,** 384-422.

LUPYAN, G. & CLARK, A. 2015. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science,* 24**,** 279-284.

MAGNUSON, J. S., MCMURRAY, B., TANENHAUS, M. K. & ASLIN, R. N. 2003. Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science,* 27**,** 285-298.

MALMIERCA, M. S. & HACKETT, T. A. 2010. Structural organization of the ascending auditory pathway. *The Auditory Brain***,** 9-41.

MATUSCHEK, H., KLIEGL, R., VASISHTH, S., BAAYEN, H. & BATES, D. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language,* 94**,** 305-315.

MAYE, J., WEISS, D. J. & ASLIN, R. N. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science,* 11**,** 122-134.

MAYE, J., WERKER, J. F. & GERKEN, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition,* 82**,** B101-B111.

MCCLELLAND, J. L. & ELMAN, J. L. 1986. The TRACE model of speech perception. *Cognitive psychology,* 18**,** 1-86.

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**,** 219.

MCMURRAY, B., SAMUELSON, V. M., LEE, S. H. & TOMBLIN, J. B. 2010. Individual differences in online spoken word recognition: Implications for SLI. *Cognitive psychology,* 60**,** 1-39.

MCMURRAY, B., TANENHAUS, M. K. & ASLIN, R. N. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition,* 86**,** B33-B42.

MCMURRAY, B., TANENHAUS, M. K. & ASLIN, R. N. 2009. Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of memory and language,* 60**,** 65-91.

MCQUEEN, J. M. 2003. The ghost of Christmas future: didn't scrooge learn to be good?: Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science,* 27**,** 795-799.

MCQUEEN, J. M., EISNER, F. & NORRIS, D. 2016. When brain regions talk to each other during speech processing, what are they talking about? Commentary on Gow and Olson (2015). *Language, Cognition and Neuroscience,* 31**,** 860-863.

MILLER, J. L. 1997. Internal structure of phonetic categories. *Language and cognitive processes,* 12**,** 865-870.

MILLER, J. L., GREEN, K. & SCHERMER, T. M. 1984. A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics,* 36**,** 329-337.

NORRIS, D., MCQUEEN, J. M. & CUTLER, A. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences,* 23**,** 299-325.

OLESON, J. J., CAVANAUGH, J. E., MCMURRAY, B. & BROWN, G. 2017. Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical methods in medical research,* 26**,** 2708-2725.

PEREIRA, O., GAO, Y. A. & TOSCANO, J. C. 2018. Perceptual Encoding of Natural Speech Sounds Revealed by the N1 Event-Related Potential Response. *Auditory Perception & Cognition,* 1**,** 112-130.

PISONI, D. B. & TASH, J. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & psychophysics,* 15**,** 285-290.

PORT, R. 2007. How are words stored in memory? Beyond phones and phonemes. *New ideas in psychology,* 25**,** 143-170.

RAO, R. P. & BALLARD, D. H. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience,* 2**,** 79.

ROBERTSON, E. K., JOANISSE, M. F., DESROCHES, A. S. & NG, S. 2009. Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science,* 12**,** 753-767.

SALMINEN, N. H., TIITINEN, H. & MAY, P. J. C. 2009. Modeling the categorical perception of speech sounds: A step toward biological plausibility. *Cognitive, Affective, & Behavioral Neuroscience,* 9**,** 304-313.

SEEDORFF, M., OLESON, J. & MCMURRAY, B. 2018. Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language,* 102**,** 55-67.

SEEDORFF, M., OLESON, J. & MCMURRAY, B. submitted. Maybe maximal: Good enough mixed models optimize power while controlling Type I error.

SHARMA, A., MARSH, C. M. & DORMAN, M. F. 2000. Relationship between N 1 evoked potential morphology and the perception of voicing. *The Journal of the Acoustical Society of America,* 108**,** 3030-3035.

SZOSTAK, C. M. & PITT, M. A. 2013. The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics,* 75**,** 1533-1546.

TANENHAUS, M. K., SPIVEY-KNOWLTON, M. J., EBERHARD, K. M. & SEDIVY, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science,* 268**,** 1632-1634.

TOMBLIN, J. B., RECORDS, N. L., BUCKWALTER, P., ZHANG, X., SMITH, E. & O'BRIEN, M. 1997. Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research,* 40**,** 1245-1260.

TOSCANO, J. C., MCMURRAY, B., DENNHARDT, J. & LUCK, S. J. 2010. Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science,* 21**,** 1532-1540.

VANDEWALLE, E., BOETS, B., GHESQUIERE, P. & ZINK, I. 2012. Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities,* 33**,** 635-644.

WERKER, J. F. & TEES, R. C. 1987. Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology/Revue canadienne de psychologie,* 41**,** 48.

ZELLOU, G. & DAHAN, D. 2019. Listeners maintain phonological uncertainty over time and across words: The case of vowel nasality in English. *Journal of Phonetics,* 76**,** 100910.

Supplement to:

Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations.

McCall E Sarrett[1,*]

Bob McMurray[2]

and

Efthymia C Kapnoula[2,3]

[1] Interdisciplinary Graduate Program in Neuroscience, 356 Medical Research Center, University of Iowa, Iowa City, IA, 52242

[2] Department of Psychological & Brain Sciences, W311 Seashore Hall, University of Iowa, Iowa City, IA, 52242

[3] Basque Center on Cognition, Brain, & Language, Mikeletegi Pasealekua, 69, 20009 Donostia, Gipuzkoa, Spain

**corresponding author* *McCall E Sarrett*
e-mail *mccall-sarrett@uiowa.edu*

S1: Stimulus details.

Below we quantify relevant characteristics of both the biasing sentences as well as the target words.

- Cloze probabilities were determined in a separate free-response norming experiment, run on Amazon Mechanical Turk with native English speakers. For each sentence we computed the likelihood of the predicted response. These were not significantly different between /b/-biased and /p/-biased sentence contexts (t(58) = 1.188, p = .23).
    - /b/-biased average: 0.86 ± 0.11
    - /p/-biased average: 0.82 ± 0.15

- Duration was calculated in Praat from the original, naturally-spoken utterance of the female native speaker with the coarticulation matching the sentence bias, and were not significantly different between /b/-biased and /p/-biased sentence contexts (t(58) = 0.719, p = .47).
    - /b/-biased average: 2165.4 ± 708.8 msec
    - /p/-biased average: 2280.6 ± 517.0 msec

- Idioms were determined by whether or not the target word referred to its literal definition, or whether it was part of a larger phrase with an obfuscated meaning (i.e. "the birds and the *bees*"), and were not significantly different between conditions (χ^2(1, N=60) = 0.626, p = .30).
    - /b/-biased average: 1 idiom / 30 total phrases
    - /p/-biased average: 3 idioms / 30 total phrases

- Compound words included proper nouns (i.e. Super *Bowl*), words where the target made up part of that word (i.e. i*Pad*), and also word pairs with extremely high co-occurrences (i.e. amusement *park*). The /p/-biased condition did have significantly more compound words than the /b/-biased condition (χ^2(1, N=60) = 8.369, p = .002).
    - /b/-biased average: 4 compound words / 30 total phrases
    - /p/-biased average: 16 compound words / 30 total phrases

- Word frequency was calculated for each word. A dash indicated that the word didn't show up in the corpus. We used both Brown frequency and Kucera-Francis frequency; word frequency was not significant different between conditions (Brown: t(18) = 1.394, p = .18; Kucera-Francis: t(18) = 1.343, p = .19).
    - /b/-biased average
        - Brown: 32.3 ± 69.0
        - Kucera-Francis: 167 ± 303.7
    - /p/-biased average:
        - Brown: 0.6 ± 0.84
        - Kucera-Francis: 23.7 ± 29.5

Tables S1a and S1b show the full set of experimental stimuli and their accompanying characteristics.

Table S1a: ***/B/-biased experimental stimuli.***

Sentence	*Cloze prob.*	*Duration (msec)*	*Idiom*	*Compound Word*	*Target Word*	*Brown freq.*	*K-F freq.*
She stole my doll, so I asked her to give it	*.91*	*3338.3*			*back*	221	967
Don't worry, I got your	*.88*	*1625.2*					
I can't reach to scratch my	*.82*	*1957.0*					
There are some good news, and some	*.81*	*1876.2*			*bad*	64	142
Well, that's too	*.78*	*1051.6*					
Sometimes he's good, sometimes he's	*.75*	*3198.6*					
The dog started to	*.88*	*1217.1*			*bark*	-	14
The outer part of a tree is called	*.82*	*1990.1*					
Quiet dogs sometimes also	*.79*	*1920.8*					
If you are dirty, you should get in the tub and take a	*1.0*	*3171.5*			*bath*	3	26
She took a nice warm	*.84*	*1540.8*					
The little girl took a bubble	*.76*	*1661.2*					
He enjoyed the ocean air, so he often went to the	*.96*	*3348.7*			*beach*	1	61
She lied on the sandy	*.91*	*1397.3*					
While in LA, we went to Venice	*.65*	*2213.5*					
Winnie the Pooh is a hungry	*1.0*	*1762.6*			*bear*	9	57
The white furry animal is a polar	*.96*	*2535.4*					
The animal on California's flag is a	*.75*	*2868.2*					
Honey is made by	*1.0*	*1324.6*			*bees*	1	-
I had to talk to my son about the birds and the	*1.0*	*2683.5*	X				
Out of the hive, came the	*.76*	*1792.1*					
The governor vetoed the	*1.0*	*1937.7*			*bill*	7	143
I paid my water	*.89*	*1633.2*					
In order to register, you must pay your University	*.88*	*3794.6*					
I poured my cereal into the	*.96*	*2026.6*			*bowl*	2	23
Beyoncé sang at the Super	*.96*	*1979.7*		X			
We're going to Florida to watch the Orange	*.92*	*2743.1*		X			
I have a play station and an X-	*1.0*	*2700.1*		X	*box*	15	70
She played with a jack-in-the-	*.70*	*1980.2*		X			
It's an empty jewelry	*.65*	*1693.3*					

Table S1b: */P/-biased experimental stimuli.*

Sentence	*Cloze*	*Duration*	*Idiom*	*Compound Word*	*Target Word*	*Brown freq.*	*K-F freq.*
A school is to fish what to wolves is a	*1.0*	*2894.8*			*pack*	2	25
For school, I need a back	*.90*	*2091.1*		X			
I ran out of cigarettes, so I'll go buy a	*.85*	*2462.4*					
You move the computer mouse on a mouse	*.96*	*2307.0*		X	*pad*	-	8
In the lake, we saw a lily	*.95*	*1852.0*		X			
The iPhone is smaller than an i	*.89*	*2598.3*		X			
There are several roller coasters in that amusement	*1.0*	*2843.5*		X	*park*	2	94
Driving in Iowa City is miserable because there's never anywhere to	*.90*	*3713.2*					
In New York, there is the Central	*.72*	*1958.7*		X			
She was led down the garden	*.86*	*1675.5*			*path*	-	44
Unfortunately, our shop is a bit off the beaten	*.59*	*3031.4*	X				
He went down the wrong	*.39*	*1449.1*					
Super Mario found Princess	*1.0*	*2070.7*		X	*peach*	-	3
The state fruit of Georgia is the	*.87*	*2041.3*					
Isn't she just a Georgia	*.68*	*1729.5*	X				
My running shoes are done, I need to get me a new	*.88*	*2967.6*			*pair*	1	6
These gloves are on sale, I'll get a	*.70*	*2621.9*					
Another word for "couple" is a	*.67*	*2033.3*					
Fergie is in the Black-Eyed	*.95*	*2190.5*		X	*peas*	-	-
Hummus is made out of chick	*.83*	*1608.8*		X			
I'm allergic to sugar snap	*.62*	*1812.3*		X			
Relax and take a chill	*1.0*	*2038.6*	X		*pill*	-	15
I order to sleep I take a sleeping	*.90*	*2420.2*					
If you are in pain, take a pain	*.57*	*2453.6*					
Santa lives at the North	*1.0*	*1725.4*		X	*pole*	-	18
There are no penguins in the South	*.70*	*2195.0*		X			
Greenland is right under the North	*.81*	*2074.8*		X			
I got a vaccination for the chicken	*.92*	*2494.5*		X	*pox*	1	1
A lot of Native Americans died of small	*.91*	*3058.1*		X			
He is down with chicken	*.74*	*2005.0*		X			

Table S1c. ***All filler stimuli.***

B-biased			*P-biased*		
Sentence	*Cloze*	*Target Word*	*Sentence*	*Cloze*	*Target Word*
A pub is just another name for a	*1.0*	*bar*	Miss Universe wished for world	*.95*	*peace*
I'll just get something from the salad	*.96*		Speak up, or forever hold your	*.95*	
If you want a drink, you should go to a	*.81*		War is the opposite of	*.92*	
Whatever floats your	*.86*	*boat*	The head of the Catholic church is the	*.84*	*pope*
We went fishing on my dad's	*.71*		In the Vatican, they elected the new	*.82*	
I am bored of sailing, I want a motor	*.68*		He wore a tall white hat and a big cross, just like the	*.68*	
Every Thanksgiving, my sister and I fight over the wish	*1.0*	*bone*	I need a pencil or a	*.96*	*pen*
I've sprained my ankle, but I've never broken a	*1.0*		There is no more ink in this	*.86*	
The dog ran out to bury her	*1.0*		The pitcher just went into the bull	*.35*	
On Kindle, you can read an e-	*.73*	*book*	I need an Advil because I'm in	*.91*	*pain*
He's so predictable, like an open	*.95*		I'm completely numb, I feel no	*.83*	
Let me read my	*.52*		You just need to step up to the	*.83*	*plate*
My feet stayed dry because I wore my rain	*.95*	*boots*	The walls need another coat of	*.96*	*paint*
This dress goes with these cowboy	*.91*		They're as close as two coats of	*.63*	
It's snowing, so put on a pair of snow	*.82*		She made a list with cons and	*1.0*	*pros*

S2: Effect of Coarticulation

Coarticulation in the final phoneme(s) of the carrier sentence was manipulated orthogonally to the other factors as it provides a secondary cue to voicing (the primary cue being Voice Onset Time, or VOT). Previous psycholinguistic work has shown that coarticulatory information of a preceding phoneme influences how listeners categorize phonemes (Holt, Lotto, & Kluender, 2000; Mann & Repp, 1980). While this was not a variable of core interest (rather, a methodological counterbalancing factor), we conducted exploratory analyses which included coarticulation as a factor in our analysis of the response data.

Figure S2 shows phoneme response curves as a function of coarticulatory condition. The model is the same as described in Figure 1 (left panel, main text) and is further detailed in S3 below. Coarticulation shifted phoneme responses in the predicted direction (β = 0.37, SE = 0.022, z = 17.07, p < .0001).

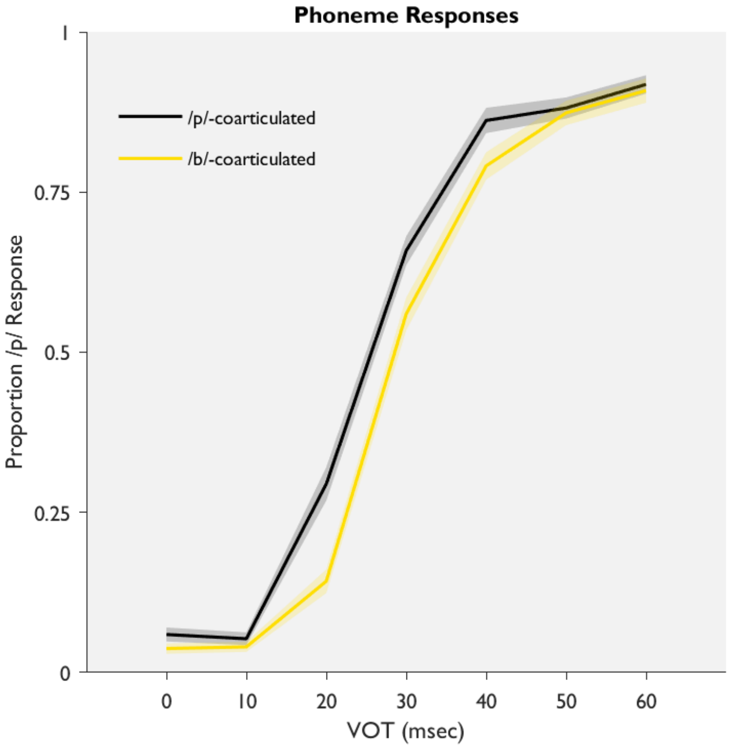

Figure S2. Phoneme categorization curves split by coarticulation of the sentence-final phoneme (N = 31, p < .0001). Shaded region is standard error of the mean.

S3: Model descriptions & detailed output for phoneme responses and reaction times.

Phoneme response categorization data.

Response categorization was analyzed with a mixed effects model with the binomial link function. The fixed effects are Bias and Coarticulation as factors and VOT as a continuous predictors, and the random effects (chosen through forward model selection) are a random intercept of Subject, as well as random slopes of Bias and VOT on subject, random intercept of Item and a random slope of Bias on Item. Fixed effects are coded using the scheme described in Table S3a.

The formula for the model (in LMER) notation is provided in (1).

```
results <- glmer( Response ~ Bias + VOT + Coarticulation + ( 1 + Bias + VOT | Subject )     (1)
… + ( 1 + Bias | Item ), data = currentdata, family = binomial )
```

Full model output for fixed effects is shown in Table S3b.

Table S3a. Fixed effects coding scheme for response categorization LME.

Bias		VOT	
actual	code	actual	code
B	-1	0	-1
P	1	10	-0.667
		20	-0.333
Coarticulation		30	0
actual	code	40	0.333
B	-1	50	0.667
P	1	60	1

Table S3b. Response categorization LME output of fixed effects.

Fixed Effects	B	SE	z	*p*	
(Intercept)	0.50	0.29	1.73	.08	
Bias	0.47	0.13	3.48	.0005	***
VOT	5.36	0.30	18.15	< .0001	***
Coarticulation	0.37	0.02	17.07	< .0001	***

Reaction time data.

Reaction time data (RT) was also analyzed with a mixed effects model. RTs were transformed using natural log. The fixed effects are Bias, included as a factor, VOT, included as a continuous predictor, and VOT^2, also included as a continuous predictor. The coding scheme is shown in Table S3c. The random effects (chosen through forward model selection) are a random intercept of Subject, random slopes of Bias, VOT, and VOT^2, random intercept of Item, as well as random slopes of Bias, VOT, and VOT^2 by Item.

The formula for the model (in LMER) notation is provided in (2).

results <- lmer(RT_log ~ Bias*(VOT + VOT^2) + (1 + Bias + VOT + VOT^2) || Item) +
(1 + Bias + VOT + VOT^2) | Subject), data = currentdata) (2)

Full model output for fixed effects is shown in Table S3d.

Table S3c. Fixed effects coding scheme for RT LME..

Bias		VOT		VOT^2	
actual	**code**	**actual**	**code**	**actual**	**code**
B	-1	0	-1	0	1
P	1	10	-0.667	10	-0.111
		20	-0.333	20	-0.778
		30	0	30	-1
		40	0.333	40	-0.778
		50	0.667	50	-0.111
		60	1	60	1

Table S3d. RT LME output of fixed effects.

Fixed Effects	B	SE	df	t	*p*	
(Intercept)	2.44	.03	34.08	87.44	< .001	***
Bias	-.0068	.005	10.19	-1.37	.19	
VOT	-.029	.009	20.99	-3.01	.007	**
VOT^2	-.039	.005	22.74	-7.36	< .001	***
Bias x VOT	.0006	.003	23370	.19	.85	
Bias x VOT^2	.0068	.003	23370	2.57	.01	*

Note: Degrees of freedom are estimated using the Satterthwaite approximation, as implemented in the lmerTest package in R. This is a standard approach for calculating degrees of freedom in linear mixed effects models. Using these approximations has not been shown to result in higher Type I error rates (see Seedorff et al., submitted).

S4: Detailed results of LME analysis of EEG over time.

We ran an LME analysis of the EEG signal every 2 msec to determine when different factors significantly predicted the averaged scalp voltage from frontocentral electrodes (Fz, F3, F4, Cz, C3, C4). This set of electrodes was chosen because they have been implicated in acoustic processing at the canonical N1 (Fz, F3, F4; Toscano, McMurray, Dennhardt, & Luck, 2010) as well as semantic processing in the auditory modality at the canonical N400 (Cz, C3, C4; Kutas & Federmeier, 2011).

Fixed effects are Bias and Coarticulation, included as factors, and linear VOT and quadratic VOT as continuous predictors. Table S4a shows the coding scheme used. Alpha levels for each factor were corrected for multiple comparisons and are shown in Table S4b (Oleson, Cavanaugh, McMurray, & Brown, 2017; Seedorff, Oleson, & McMurray, 2018).

Random effects structure are a random intercept of Subject, random intercept of Item, and random slope of Bias on Item. Random effects structure was determined by choosing representative timepoints along the epoch, and testing different models at said timepoints. The model with the lowest Akaike's Information Criterion (AIC) at the majority of timepoints was then selected to run across the full epoch.

The formula for the model (in LMER) notation is provided in (3).

$$\begin{aligned}&\text{results <- lmer(Voltage} \sim \text{Bias * (VOT + VOT}^{2}\text{) + Coarticulation + (1 | Subject) +}\\&\qquad\qquad\text{(1 + Bias || Item), data = currentdata)}\end{aligned}\tag{3}$$

Table S4a. Fixed effects coding scheme for EEG LME over time.

Bias		**VOT**		**VOT²**	
actual	**code**	**actual**	**code**	**actual**	**code**
B	-1	0	-1	0	1
P	1	10	-0.667	10	-0.111
		20	-0.333	20	-0.778
Coarticulation		30	0	30	-1
actual	**code**	40	0.333	40	-0.778
B	-1	50	0.667	50	-0.111
P	1	60	1	60	1

Table S4b. Error-corrected alphas for predictors (main effects and interactions).

	Bias	**Coarticulation**	**VOT**	**VOT²**	**Bias x VOT**	**Bias x VOT²**
FWEC α's	0.0169	0.0168	0.0227	0.0206	0.0259	0.0193

Finally, Table S4c shows a summary of the LME outputs, shown in Figure 3 (main text), with the time windows (a) of all significant predictors after correction for multiple comparisons. The last column in Table S4c gives the formula with beta coefficients for the midpoint of the time window, corresponding to the time shown in (b). Select timepoints from later in the epoch are plotted in Figure S4, which were not included in the main text.

Table S4c. Significant time windows (a) and corresponding formulas with beta weights from timepoint (b).

Time (msec)		**Significant**	**Formula to calculate μV**
(a)	**(b)**	**Predictor**	$B_0 + B_1X_1 + B_2X_2 + \ldots$
96 – 118	108	Coarticulation	1.289 + (0.225*Coarticulation)
80 – 164	122	VOT^2	1.371 + (0.103*VOT) + (-0.407*VOT^2)
130 – 220	176	VOT	-0.773 + (0.939*VOT)
208 – 270	238	Bias x VOT^2	0.825 + (0.043*Bias) + (0.079*VOT) + (-0.292*VOT^2) + (0.347*Bias*VOT) + (-0.334*Bias*VOT^2)
232 – 274	254	VOT^2	0.840 + (0.151*VOT) + (-0.301*VOT^2)
228 – 352	290	Bias x VOT	0.332 + (0.002*Bias) + (0.515*VOT) + (0.712*Bias*VOT)
260 – 340	300	VOT	0.130 + (0.620*VOT)
276 – 292	284	Coarticulation	0.441 + (-0.190*Coarticulation)
372 – 416	394	VOT^2	0.343 + (-0.150*VOT) + (0.211*VOT^2)
388 – 410	398	Bias x VOT^2	0.375 + (0.066*Bias) + (-0.170*VOT) + (0.212*VOT^2) + (-0.515*Bias*VOT) + (0.216*Bias*VOT^2)
378 – 486	432	Bias x VOT	0.503 + (0.032*Bias) + (-0.163*VOT) + (-0.758*Bias*VOT)
468 – 492	478	Coarticulation	-0.007 + (0.188*Coarticulation)
470 – 498	484	Bias x VOT^2	-0.060 + (0.073*Bias) + (0.080*VOT) + (0.122*VOT^2) + (-0.263*Bias*VOT) + (-0.326*Bias*VOT^2)
506 – 528	518	Bias	-0.184 + (0.228*Bias)
520 – 542	530	Bias x VOT^2	-0.232 + (0.132*Bias) + (0.127*VOT) + (-0.033*VOT^2) + (0.211*Bias*VOT) + (-0.247*Bias*VOT^2)
530 – 580	556	Bias x VOT	-0.531 + (0.082*Bias) + (-0.004*VOT) + (0.329*Bias*VOT)
670 – 762	716	VOT	-0.371 + (-0.294*VOT)
706 – 782	744	VOT^2	0.307 + (-0.274*VOT) + (-0.266*VOT^2)
804 – 848	826	VOT	0.899 + (0.362*VOT)
864 – 882	874	VOT	0.517 + (0.238*VOT)

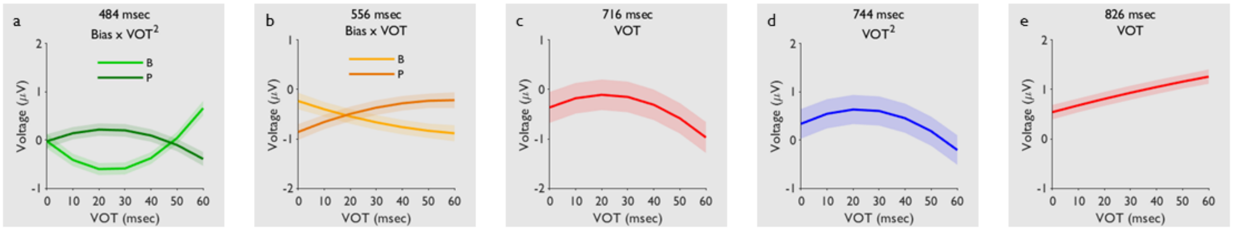

Figure S4. Effects taken from peak effect timepoints show change in predicted voltage as a function of different predictors, calculated from a parametric bootstrap on the estimated model that time. This bootstrap estimated the predicted voltage for a "new" subject. Individual effects (like VOT) were set to the original values used in the model; for effects not shown, corresponding IVs were set to 0. Standard error of the model's predicted value is shown by the shaded region. In the calculations for B-F, the terms for VOT and VOT^2 move together, as they reflect different polynomial transformations of the same variable. VOT^2 shows the effect of phonemic ambiguity (D). VOT is acoustic cue encoding (C and E). Bias × VOT^2 shows the differential effect of predictions from the sentence Bias depending on whether the incoming VOT is near category boundary (ambiguous) or not (A). Bias × VOT is the integration of semantic/contextual information with the incoming spoken word (B).

S5: Replication of LME analysis over time at centroparietal electrodes.

We chose frontocentral electrodes for our main analysis to capture electrodes sensitive both to auditory processing as well as semantic/contextual processing (see S4). However, semantic/contextual integration indicated by the N400 is often strong over parietal electrodes as well. Thus, to be sure that our analyses are robust and not particular to that subset of electrodes, we re-ran the model shown in (3) using centroparietal electrodes (Cz, C3, C4, Pz, P3, P4) to capture both a potentially more parietal N400 and centrally located auditory information. Results are below and can be directly compared to Figure 3 in the main text.

We see again an early and long-lasting effect of VOT (from around 130 msec to ~350 msec) that re-emerges around 660 msec. As in the primary analysis, there was little overall main effect of Bias, but a Bias × VOT interaction from around 210 msec to 500 msec (and later). Crucially, the Bias × VOT^2 interaction was significant in the 210 to ~280 msec range, as was also observed in the primary analyses at frontocentral electrodes. Thus, none of the conclusions—particularly those reflecting context effects—are affected by which of these electrode subsets were included in the analysis.

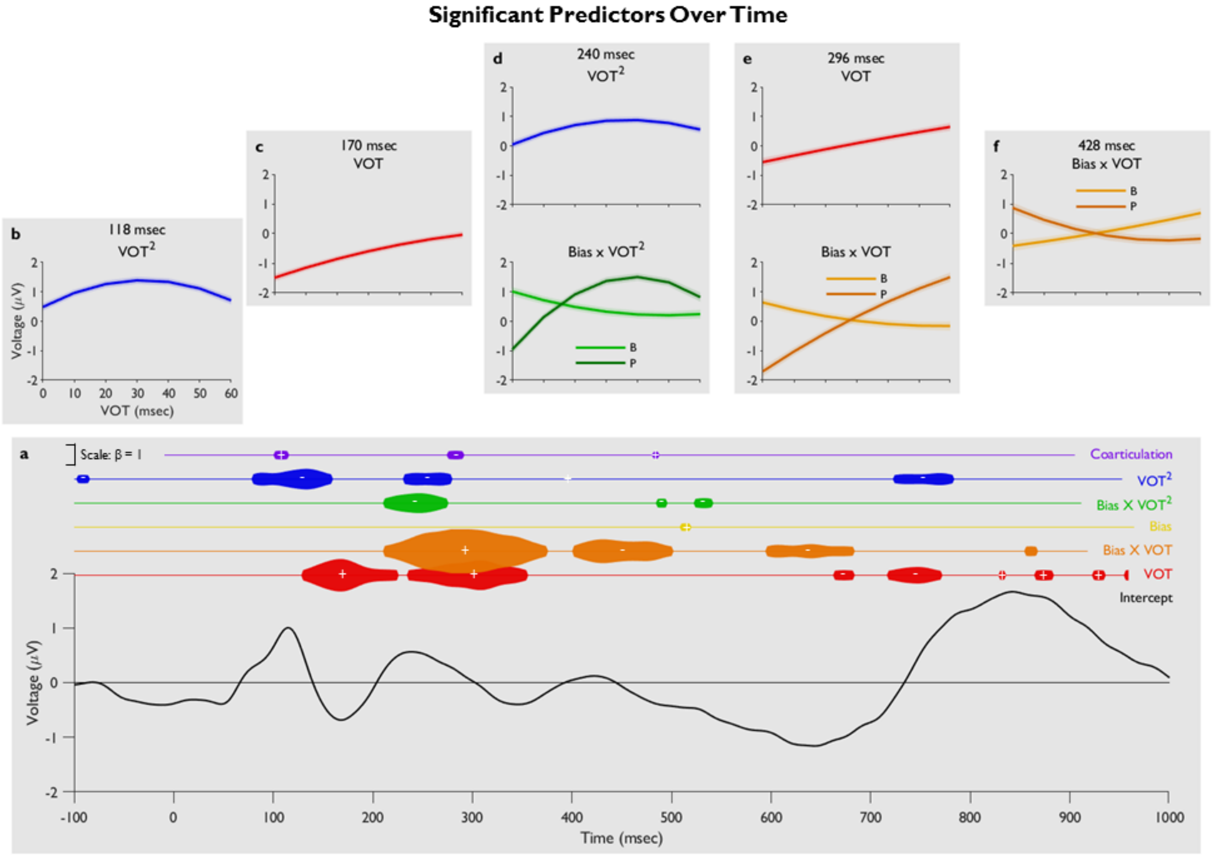

Figure S5. Replication of LME over time at centroparietal electrodes (C3, Cz, C4, P3, Pz, P4). (A) The full model output over time. The intercept (estimated voltage from the model output) is shown on the bottom line graph with beta coefficients of significant predictors shown above; the width of the balloon corresponds to the strength of the effect and the positive/negative symbol corresponds to the direction of the effect. (B-F) Effects taken from peak effect timepoints show change in predicted voltage as a function of different predictors, calculated from a parametric bootstrap on the estimated model that time. This bootstrap estimated the predicted voltage for a "new" subject. Individual effects (like VOT) were set to the original values used in the model; for effects not shown, corresponding IVs were set to 0. Standard error of the model's predicted value is shown by the shaded region. In the calculations for B-F, the terms for VOT and VOT^2 move together, as they reflect different polynomial transformations of the same variable. VOT^2 shows the effect of phonemic ambiguity (A and D). VOT is acoustic cue encoding (C and E). Bias × VOT^2 shows the differential effect of predictions from the sentence Bias depending on whether the incoming VOT is near category boundary (ambiguous) or not (D). Bias × VOT is the integration of semantic/contextual information with the incoming spoken word (E and F).

Table S5. Significant time windows (a) and corresponding formulas with beta weights from timepoint (b) for LME at centroparietal electrodes.

Time (msec)		**Significant Predictor**	**Formula to calculate µV**
(a)	**(b)**		$B_0 + B_1X_1 + B_2X_2 + \ldots$
80 - 158	120	VOT^2	0.94088 + (0.133*VOT) + (-0.408*VOT^2)
102 – 114	108	Coarticulation	0.902 + (0.201*Coarticulation)
130 – 224	178	VOT	-0.609 + (0.642*VOT)
212 – 274	244	Bias x VOT^2	0.555 + (0.093*Bias) + (0.276*VOT) + (-0.281*VOT^2) + (0.706*Bias*VOT) + (-0.450*Bias*VOT^2)
212 – 374	294	Bias x VOT	0.094 + (-0.053*Bias) + (0.585*VOT) + (1.008*Bias*VOT)
232 – 278	236	VOT^2	0.562 + (0.235*VOT) + (-0.256*VOT^2)
236 – 354	296	VOT	0.071 + (0.601*VOT)
276 – 290	284	Coarticulation	0.221 + (-0.146*Coarticulation)
402 – 500	450	Bias x VOT	-0.069 + (-0.062*Bias) + (-0.061*VOT) + (-0.461*Bias*VOT)
482 – 486	484	Coarticulation	-0.385 + (0.158*Coarticulation)
486 – 494	490	Bias x VOT^2	-0.410 + (0.052*Bias) + (-0.063*VOT) + (0.179*VOT^2) + (-0.323*Bias*VOT) + (-0.205*Bias*VOT^2)
510 – 518	514	Bias	-0.459 + (0.175*Bias)
524 – 540	532	Bias x VOT^2	-0.520 + (0.052*Bias) + (0.125*VOT) + (0.129*VOT^2) + (0.009*Bias*VOT) + (-0.228*Bias*VOT^2)
596 – 682	640	Bias x VOT	-1.156 + (-0.095*Bias) + (-0.068*VOT) + (-0.323*Bias*VOT)
664 – 682	674	VOT	-0.993 + (-0.230*VOT)
718 – 770	744	VOT	0.337 + (-0.374*VOT)
724 – 782	754	VOT^2	0.649 + (-0.356*VOT) + (-0.227*VOT^2)
830 – 834	832	VOT	1.620 + (0.206*VOT)
856 – 866	862	Bias x VOT	1.593 + (0.034*Bias) + (0.186*VOT) + (0.221*Bias*VOT)
866 – 882	872	VOT	1.570 + (0.236*VOT)
924 – 934	930	VOT	0.798 + (0.228*VOT)
926 – 960	942	Bias x VOT	0.637 + (0.038*Bias) + (1.881*VOT) + (0.301*Bias*VOT)
956 – 968	968	VOT	0.369 + (0.215*VOT)

S6: Effect of Phonemic Ambiguity on EEG waveform.

Phonemic ambiguity was coded as distance from the middle VOT (30 msec) in our continuum and included as the term VOT^2 in our LME model for predicting scalp voltage, detailed in Table S4a. Figure S6 shows the waveform at frontocentral electrodes split by steps from the most ambiguous token. The most visible effect can be seen at the canonical P50 and P200 peaks, where the most ambiguous token yields the highest voltage, roughly 110 msec and 250 msec, respectively.

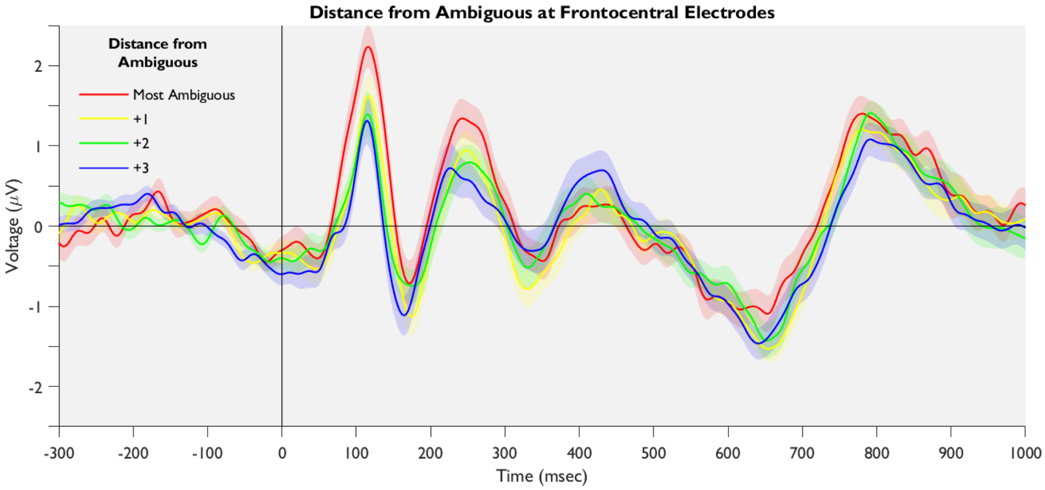

Figure S6. Averaged EEG waveform as a function of distance from the most ambiguous VOT, timelocked to the presentation of the target word (N = 31).

S7: Scalp Topographies.

Our primary analyses focused at frontocentral electrodes, where acoustic (N1) and semantic (N400) information can be recorded. However, full scalp topographies are useful for a variety of reasons, and as such, are reported below.

The canonical P1/P50 can be best seen at 100 to 150 msec; the N1 at 150 to 200 msec; the P2 from 200 to 300 msec; N400 starting at 400 msec and peaking at 700 msec; and finally we see some visual evoked activity later in the epoch after ~800 msec after word offset, when response options are presented on the screen.

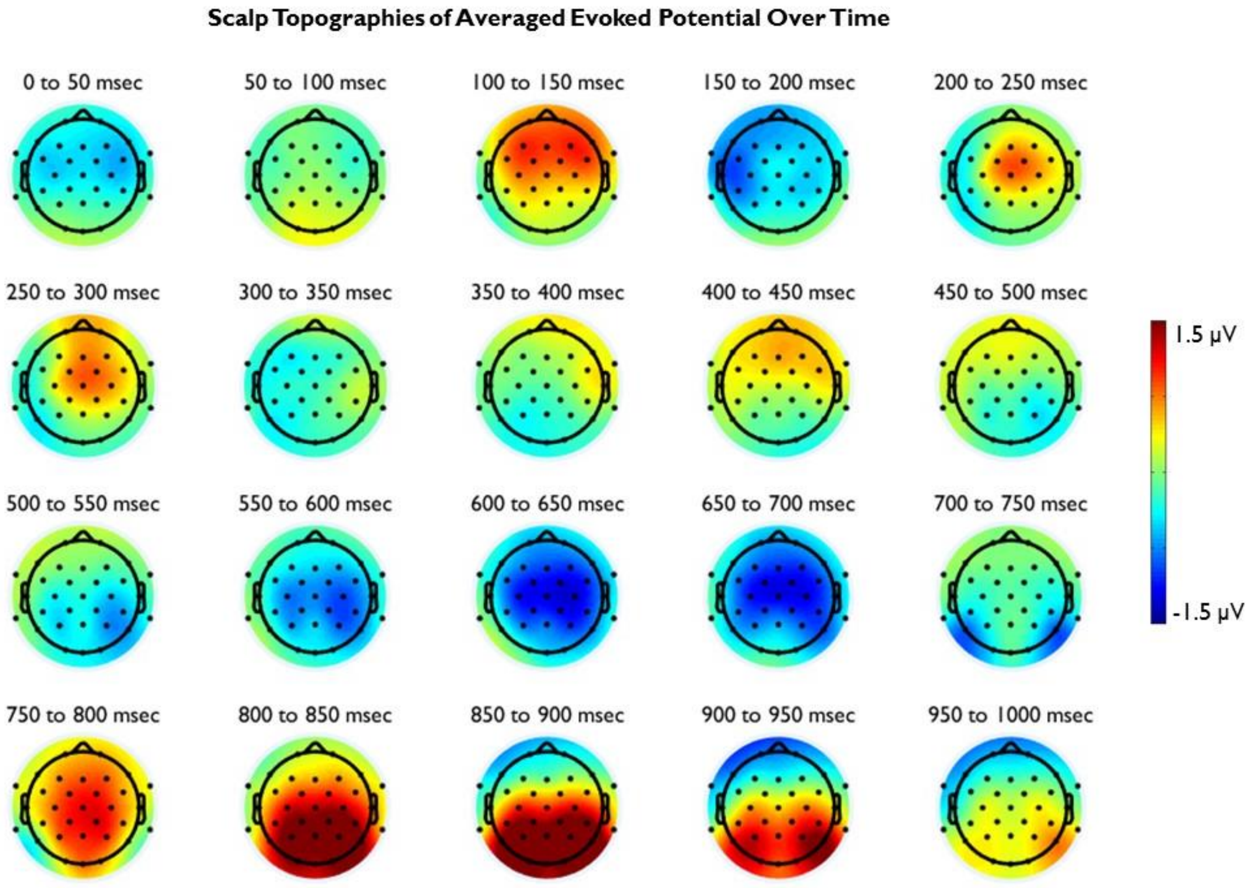

Figure S7. Scalp topography of grand average EEG waveform, averaged in 50 msec bins post-target word onset (N = 31).

References.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America, 108*(2), 710-722.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics, 28*(3), 213-228.

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical methods in medical research, 26*(6), 2708-2725.

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of memory and language, 102*, 55-67.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21*(10), 1532-1540.