Running Head: ERPS OF LEXICAL BIAS

ACCEPTED at Cognition, 12/10/19

Early lexical influences on sublexical processing in speech perception: Evidence from electrophysiology

Colin Michael Noe* & Simon Fischer-Baum

Department of Psychological Sciences, Rice University, Houston, TX

*Corresponding author

Department of Psychological Sciences, MS-25
Rice University
P.O. Box 1891
Houston, TX 77251

Tel: 281-723-5687
colinmichaelnoe@gmail.com

Authors note: Research reported in this publication was supported by the National Science Foundation under grant numbers 1250104 and 1752751

Abstract

Contextual information influences how we perceive speech, but it remains unclear at which level of processing contextual information merges with acoustic information. Theories differ on whether early stages of speech processing, like sublexical processing during which articulatory features and portions of speech sounds are identified, are strictly feed-forward or are influenced by semantic and lexical context. In the current study, we investigate the time-course of lexical context effects on judgments about the individual sounds we perceive by recording electroencephalography as an online measure of speech processing while subjects engage in a lexically biasing phoneme categorization task. We find that lexical context modulates the amplitude of the N100, an ERP component linked with sublexical processes in speech perception. We demonstrate that these results can be modeled in an interactive speech perception model and are not well fit by any established feed-forward mechanisms of lexical bias. These results support interactive speech perception theories over feed-forward theories in which sublexical speech perception processes are only driven by bottom-up information.

Keywords: speech perception, N100 ERP, Ganong effect, TRACE, feedback

1. Introduction

A major question that cuts across different domains in the cognitive sciences is how our perception of bottom-up sensory inputs is shaped by top-down knowledge about what we are expecting to perceive (e.g., Rumelhart, 1976; Pylyshyn, 1984; Chater & Oaksford, 1990; McClelland et al., 2014). The current study looks at this question in the context of speech perception. There is a clear consensus that both acoustic, bottom-up information, and top-down knowledge about words and their meanings influence which speech sounds we perceive. But there is considerable disagreement about the processing stage at which integration of bottom-up and top-down information occurs.

In speech perception, debate has focused on *sublexical* stages of speech perception, in which speech sound representations (e.g., phonemes) are accessed from the acoustic input. The major question has been whether this sublexical stage is independent from or influenced by top-down information. Top-down information in speech perception takes multiple forms, including lexical knowledge about which speech sound sequences form words, semantic knowledge about which words fit in a given context, or syntactic knowledge about what fits grammatically into the sentence or phrase. *Interactive* speech perception theories argue that, during the course of speech perception, all levels of processing are informed by and constrained by higher-level knowledge (e.g., TRACE: McClelland & Elman, 1986; C-Cure: McMurray & Jongman, 2011; and TISK: Hanngan, Magnuson, & Grainger, 2013; see Magnuson, Mirman, Luthra, Harris & Strauss, 2018, for discussion). In interactive theories, early sublexical representations are activated both by bottom-up connections from acoustic processing and by top-down input via feedback connections from lexical and semantic processing

levels. In contrast, *feed-forward* speech perception theories argue that speech perception relies on unbiased encoding of sublexical information (Fuzzy Logic Model of Perception, Oden & Massaro, 1978; Cohort, Marslen-Wilson, 1987; Cairns et al., 1995, 1997; Norris, 1994; Merge, Norris, McQueen, Cutler, 2000, 2003, 2008; Ideal Listener, Kleinschmidt & Jaeger, 2015, see Norris, McQueen, & Cutler, 2016 for discussion). In feed-forward theories, early encoding of the bottom-up sensory information is independent from top-down information being processed at lexical and semantic levels. We present an experiment that uses event-related-potentials (ERPs) to measure the sublexical processing at the heart of the feedback debate in speech perception while manipulating top-down information, here, the lexical context in which the sound appears, to test for influences of lexical context on sublexical processing.

Much of the debate around feedback to sublexical levels has centered on shifts in how a speech sound is perceived as a function of the context in which it appears. These top-down effects on perception occur when a higher-level context variable influences speech sound perception. Top-down effects on speech perception have been found with many different manipulations, for example by replacing speech sounds with noise in different types of spoken stimuli (e.g., Warren, 1970; Samuel, 1981) or by varying what is semantically (e.g., Sivonen, Maess, Lattner, & Friederici, 2006; Samuel, 1981; Connine & Clifton, 1987; Groppe et al., 2010) or syntactically (e.g., Fox & Blumstein, 2016) predicted by the sentence context. Top-down effects have been observed with many different measures of perception, including judgments about which speech sounds are perceived, the rate at which phonemes are recognized, or how the eyes move between competing word candidates.

Some of the best-studied top-down effects are lexical effects on phoneme category judgments. Ganong (1980) demonstrated effects of lexical bias on phoneme identification, specifically that identification is biased towards phonemes which form familiar words (see also Burton, Baum, & Blumstein 1989; Burton & Blumstein, 1995; Connine, Clifton, & Cutler, 1987; Gow, Segawha, Alfhors, & Lin, 2008; Gow & Olson, 2015; Kingston, Levy, Rysling, & Staub, 2016; Myers & Blumstein, 2007; Newman, Sawusch, & Luce, 1997; Pitt, 1995; Pitt & Samuel, 1993, 1995). For example, if a participant is presented with a sound that is ambiguous between /d/ and /t/, they are more likely to perceive it as a /d/ with a rime like /eɪt/ since *date* is a word but **tate* is not. But they would be more likely to perceive the exact same sound as a /t/ with a rime like /eɪp/ since *tape* is a word but **dape* is not.

Lexical bias effects are readily fit by interactive models that include feedback connections (e.g., Elman & McClelland 1988; Mirman, McClelland, & Holt, 2006). In theories with feedback connections from the lexicon to sublexical levels, top-down effects such as lexical biases emerge naturally from the architecture because available information from the lexicon or contextual environment will modulate activation at the sublexical level and therefore change how incoming acoustic information is encoded and then subsequently perceived. The compatability of interactive feedback models with lexical bias has been extensively demonstrated in simulation studies (Elman & McClelland, 1988, Strauss, Harris, & Magnuson 2007, see also Section 5 of this paper). However, the necessity of feedback connections to be able to model lexical bias and other top-down effects has been called into question.

Norris, Cutler, and McQueen (2000, 2003, 2008, 2016) have demonstrated that interactivity is not necessary to explain top-down effects on phoneme judgments, such as the lexical bias observed by Ganong (1980). Specifically, lexical bias effects can be generated by an architecture without feedback, but which includes a "decision module" which acts as a late integrator of lexical and sublexical information, merging the two information sources only after sublexical processing has occurred (see R. Fox, 1984; but see also, N. Fox & Blumstein 2016). The post-perceptual merger account of lexical bias demonstrates that feedback is not necessary to explain top-down effects observed in behavior. Norris et al. (2003, 2008, 2016) have also claimed that feedback results in non-optimal performance of a speech perception system (e.g., Fraunfelder & Peeters, 1998; though see also McClelland, Mirman, Bolger, & Khaitan, 2014; McClelland, Mirman, Luthra, Harris, & Strauss, 2018).

Given the equivocal status of debates over efficiency and previous evidence, more empirical work to determine when top-down and bottom-up information are integrated during speech sound processing will be necessary. With respect to the debate around lexical bias effects specifically, the issue with much of the previous literature is that it has relied on measures with inadequate online temporal resolution to determine whether lexical bias is due to feedback or due to feed-forward post-perceptual integration of bottom-up and top-down information (Fox & Blumstein, 2016). What is needed to distinguish between interactive and feed-forward explanations is a measure of sublexical processing that can quantify sublexical representation in an online manner so that bias can be tested for within the time-course of speech sound processing. Such a measure could valuably contribute to the feedback debate because

interactive and feed-forward accounts differ in when and at what stages of processing top-down effects are predicted to occur. If an online measure of sublexical processing early in the response to speech sounds can be identified and shown to vary as a function of top-down factors like lexicality, that would provide strong evidence in favor of interactive activation models. In contrast, if the online measure of sublexical encoding is not influenced by top-down factors, this would support feed-forward theories.

To this end, we utilize an online electrophysiological signature of the early perceptual encoding of incoming speech sound features, the N100 event related potential (ERP) signal (Toscano, McMurray, Dennhardt, & Luck, 2010), to measure top-down influences on early stages of speech perception with higher temporal resolution. The N100 ERP response to speech sounds, measured with scalp-recorded electroencephalography (EEG), has been linked with sublexical processing; Toscano and colleagues (2010) found that the amplitude of the N100 is linearly related to voice-onset-time in a voicing continuum (such as a continuum from /d/ to /t/). Voice-onset-time (VOT) refers to the latency, in milliseconds, relative to a consonant onset that the vocal cords begin vibrating for the following vowel. VOT is used to distinguish voiced and unvoiced minimal pairs like /d/ from /t/. The N100-VOT relationship suggests that the N100 provides a direct window into the neural processing of phonetic features.

Strengthening the argument that the signal we are measuring in the N100 reflects sublexical processing, electrocorticography (ECoG) has also linked neural activation in the 75-175 msec time range with processing of phonetic features (Steinschneider, Schroeder, Arezzo & Vaughan, 1991; Steinschneider, Nourski, & Fishman, 2013; Mesgarani, Cheung, Johnson, & Chang, 2014; Hullet, Hamilton,

Mesgarani, Schreiner, & Chang, 2016). Taking the ERP and ECoG evidence together, sublexical encoding of phonetic features is occurring around 100 msec after the onset of a phoneme and the N100 appears to directly measure some aspects of the encoding of the sublexical feature of speech sound voicing. If the amplitude of the N100 ERP component is also sensitive to top-down information, then this suggests top-down information is influencing early sublexical stages of speech perception.

A recent study by Getz and Toscano (2019) suggests that N100 amplitude is influenced by another source of top-down influence, predictions generated from strong lexical associations. For example, in the two-word utterance “Eiffel Tower”, the word “tower” is strongly predicted after the word “Eiffel” is spoken. When presented with a stimulus with an ambiguous /dt/ spliced into the onset of “tower”, Getz and Toscano found that the N100 amplitude was significantly more /t/-like than when the ambiguous /dt/ was embedded in an opposite experimental condition (e.g., “Barbie Doll”), in which a /d/ was predicted.

In the present experiment, we use the N100 to distinguish between feed-forward and interactive accounts of Ganong lexical bias effects. Specifically, we measured the N100 ERP waveform in the context of a lexically biasing phoneme categorization task. EEG signals were continuously recorded while participants made a 2-alternative-forced-choice decision in response to voice-onset-time continua while lexical bias direction was manipulated by changing the rhyme portion following the initial phoneme. For example, participants made /d/-/t/ judgments on Date-*Tate and *Dape-Tape continua. We tested for effects of lexical bias on the amplitude of the N100. Interactive theories predict that

the N100 should be influenced by both bottom-up and top-down information; feed-forward theories predict the N100 should not be influenced by top-down information.

2. Materials and Methods

2.1 Participants. Twenty-four participants (13 female, avg. age = 19.54 yrs.) were recruited from the Rice University student population and participated under the approval of the Rice University IRB. Participants were given course credit in exchange for participation. All subjects reported normal hearing, no history of neurological disorder, no recent drug use, and learning English as a first language. All twenty-four subjects displayed normal categorization, which we defined based on previous behavioral pilot experiments as classifying >80% unvoiced at the unvoiced end of the continuum and <20% unvoiced at the voiced end of the continuum (Newman, Sawusch, & Luce, 1997). Three subjects were excluded due to a high number of artifact contaminated trials in the EEG data (>40%). One additional subject was excluded due to coding errors in the trial labels sent from the stimulus presentation computer to the EEG computer.

Sample size selection was slightly larger than comparable cognitive linguistic N100 ERP studies, such as the sample in Toscano et al. (2010) which demonstrated a VOT effect in 17 subjects and the sample size in Schneider (2017) which demonstrated talker gender effects on the N100 in 20 subjects. A 24 subject sample exceeded these previous studies and allowed for even counterbalancing of block order. Similarly, the sample size exceeded the standard in the neuroscientific N100 literature (Hoonhorst et al., 2009: 10 subjects; Horev, Most, & Pratt, 2007: 14 subjects; Martin & Boothroyd,

1999: 10 subjects; Sharma & Dorman, 1999: 16 subjects; Zaehle, Jancke, & Meyer, 2007: 18 subjects).

2.2 Materials. Two pairs of lexically biasing voicing continua were created, yielding a /d/-/t/ lexical bias pair and /g/-/k/ lexical bias pair: Date-*Tate vs. *Dape-Tape and Gate-*Cate vs. *Gake-Cake. Continua were created by cross-splicing natural voiced tokens with natural unvoiced tokens at 5 msec intervals using the Andruski et al. (1994) cross-splicing method. The voiced and unvoiced endpoint stimuli were recorded at 44,100Hz in a soundproofed recording studio with a Shure SM-58 microphone at the Rice University Digital Media Center. The voiced and unvoiced natural endpoints to be spliced together (e.g., Dape and Tape, which were recorded separately) were matched on duration, pitch, intensity, formant trajectory and frequency, and envelope shape. Lexically opposing pairs (e.g., *Dape-Tape vs Date-*Tate) were balanced as closely as possible for these same acoustic factors. Final continua were selected after verifying normal categorical perception and then a Ganong effect in a pilot experiment, reported in Appendix B, with a separate set of participants. Stimuli were 400 msec in length on average; the stimulus was embedded in a 600 msec sound file, with the plosive burst for the critical phoneme occurring at 100 msec into the file. Further acoustic details for the stimuli are listed in Appendix A.

After splicing, each word-non-word pair yielded a 9-step voice-onset-time continuum with VOT ranging from 5 msec (clearly voiced) to 45 msec (clearly unvoiced) in 5 msec increments. The VOT step (9) x Bias (2) x Place (2) yielded 36 unique critical stimuli. Each stimulus was presented 64 times for a total of 2304 critical trials per subject.

2.3 Procedure. Subjects entered the testing room and informed consent was obtained. The 32-channel ActiCap system (BrainVision Systems, Morrisville, NC) was applied with active electrode gel to an impedance of less than 25 kOhm. Subjects were then fitted with EEG-compatible headphones for stimulus presentation (ER-3, Etymotic Research, Elk Grove, IL). Subjects began with a 36-trial practice block to familiarize themselves to the task and stimuli. On each trial, subjects heard one stimulus from one continuum and were asked to choose which endpoint of the continuum the token sounded most like, the voiced endpoint or the unvoiced endpoint. Each trial began with a 750ms fixation. From 550ms - 750ms the fixation cross was bolded to indicate imminent arrival of the sound. At 750 msec the stimulus began to play. 600 msec after the onset of the stimulus, two text strings with critical phonemes at onset (e.g., *date* and **tate*) appeared on the screen, and participants had to press the f or j button to indicate which of the two phonemes they perceived, making the 2-alternative forced choice (2-AFC) judgement. EEG was continuously recorded during the task and the exact timing of the sound onset was obtained using the StimTrak/TriggerBox (BrainVision Systems) system.

The experiment consisted of 32 approximately 2.5-minute blocks. Each block consisted of 72 samples from one continuum. Each of the nine VOT steps was sampled eight times to generate one block. Order of trials was random within a block. Each continuum was used in 8/32 blocks. Order of the blocks was counterbalanced in a Latin-square design across subjects, so that which block followed which was balanced across subjects. The task portion of the experiment, not including set-up or take-down time, was approximately 1.5 hours.

The use of blocks from one continuum and therefore one biasing direction was the key contextual manipulation of the experiment. Given that the lexically disambiguating information in our continua (e.g., for the d-t pairs, the final consonant in the rimes /eɪt/ and /eɪp/) did not occur until well after that N100 generating /d/-/t/ or /g/-/k/ phoneme, the blocking of stimuli from one continua per block allowed subjects to be aware of the lexicality resulting from each of the two possible percepts (e.g., *date* and **tate*). In each trial within a block, a subject would know, except for the first trial of the block, which continuum was being sampled for that block. Blocking allowed tacit knowledge about which percept would form a word to create lexical support for the word endpoint. Note that subjects were instructed to ignore the lexicality of the endpoints and no feedback was ever given, but Ganong lexical bias occurs automatically. Blocking by continuum also simplified response mappings so that only 2-responses were needed within a given block; if we had not blocked continua, subjects would have had to map responses for both endpoints of the four continua, yielding 8 responses to manage on each trial. Like any methodological choice, this blocked instantiation of bias represents one of many ways top-down influence can be induced; we consider its strengths and limitations in the discussion and follow-up analyses.

2.4 EEG Pre-Processing. Standard EEG pre-processing techniques were applied (Luck, 2014) using the ERPLAB toolbox (Lopez-Calderon & Luck, 2014) in EEGlab (Delorme & Makeig, 2004). Data were re-referenced to the average of the mastoids. High Pass (.1 Hz half-power) and Low Pass (40Hz half-power) IIR (slope 12Hz/dB) filters were applied to the continuous data. Trial onsets were identified using the TTL pulses generated by the StimTrak marked onto the continuous EEG waveform. 500

msec trial epochs were extracted plus a 200 msec pre-stimulus period used as a baseline correction. Each 500 msec epoch began at the onset of the phoneme of interest. Artifact contaminated epochs were rejected using three automated tools in the ERPLAB toolbox, moving window threshold, absolute threshold, and covariance blink detection. Average artifact rejection rates were 15.9%. Epochs not marked as artifacts were averaged together within each condition to yield within-subject, within-condition ERP waveforms for plotting, but the mixed models were run on the single epoch amplitudes (38,491 trials).

The time window and electrode regions were defined in a collapsed localizer method (Luck & Gaspelin, 2017) – before any experimental effects were examined. This method, which was selected *a priori,* is laid out in Luck (2014) and is recognized as a statistically robust method to select the region and time-window of interest for ERP experiments with stimuli, paradigms, or populations that do not yet have an established ERP topology and time-course. The collapsed localizer method specifies that the region and time-window should be defined by looking at the global average topography and time-course of the N100 negativity, collapsing across any experimental effects of interest. The electrode region selected was chosen because it showed a negative peak at 100 msec when the data was averaged across all conditions. At the time of its selection, the VOT and bias effects of interest had not yet been examined.

The collapsed localizer identified the left-frontal quadrant (electrodes Fp1, F3, F7, C3, FC5, and T7) as the largest spatially continuous set of N100-displaying electrodes, and the average of these six electrodes was computed as the region of interest. The collapsed localizer identified the time window [60 – 130 ms] as the period

during which the ERP had a negative value around 100 msec. Accordingly, N100 ERPs were defined as the average amplitude for each trial in the time window, 60 – 130 msec, averaged across the six left fronto-central electrodes. This time-window is similar to the time-window in the previous literature (e.g., Toscano et al., 2010; Getz & Toscano, 2019), and the statistical effects were robust to variations in the time-window as is visible in Figure 3. The electrode region, though more left lateralized than Toscano et al., 2010, is similar to previous literature in showing a left lateralized auditory evoked response for speech sound processing, especially rapid spectrotemporal aspects of speech such as VOT (e.g., Sanders & Neville, 2003; Obleser, Eulitz, & Lahiri, 2003, 2004; De Fonseca, Giraud, Badier, Chauvel, & Liegeois-Chauvel, 2005; Obleser, Roskstroth, & Eulitz, 2004; Davis, Kislyuk, Kim & Sams, 2008; Hornickel, Skoe, & Kraus, 2009; Hutschison, Blumstein, & Myers, 2008). Just as with the time-window, adding in or removing electrodes did not alter the significance of the experimental effects.

2.5 Statistical Methods. A mixed effect model approach was taken to evaluate the lexical bias effects in both the categorization responses and in the N100 data. The only difference between the modelling method used for behavior and for ERP was the linking function – with the behavioral data employing a logistic linking function since the 2-AFC behavioral task generates binary outcomes and the N100 data not requiring a linking function since the ERP data are continuous.

In both behavior and N100 data, a forward stepwise model comparison approach was used. The base model contained only the fixed effect of VOT (coded as a linear variable with nine steps from 5 – 45 msec, untransformed) and the random effect of

subject, place of articulation and trial number in block. The addition of the effect of Bias Condition (coded as voiced biasing = -1, unvoiced biasing = 1) and the Bias-VOT interaction were evaluated. To evaluate whether the bias effect was largest in the mid-range VOT where the sound is ambiguous and where Ganong (1980) observed the largest bias effects, a transformation of the VOT was necessary so that the VOT:Bias interaction term would have its largest value at the mid-range VOTs. We employed a gaussian transformation where VOT was transformed using the normal distribution to an inverted U shape ($VOT - transform = k * p(VOT \sim N(\mu = 25, \sigma = 5))$), and then multiplied the result of this transformation by a constant, k, so that the value at VOT 25 was equal to 1 ($k = 12.533$). This procedure generated a value for the *VOT-transform* which maps the *Bias:VOT-transform* interaction onto the prediction of bias being largest at the most ambiguous VOTs which can be treated like a linear effect within the model. Higher order interactions were also evaluated in the model, but none significantly improved model fit.[1]

The main research question, whether lexical bias is evident at the N100, was evaluated by testing the contribution (and direction) of the Bias-VOT interaction term. We expect a positive value for the slope estimate for the interaction term where positive implies more negative amplitudes for voiced biasing (coded as -1) stimuli and more

[1] For the model fitting procedure, in both the behavior and N100 analyses, we included the maximal random effect structure supported by the data (Barr et al., 2013). Models that included random slope models did not reliably converge, so, as a result, only the random intercept models are reported.

positive for unvoiced biasing (coded as +1) stimuli. Comparison of the interaction term versus the main effect of bias provides an estimate of how well the N100 bias matches the VOT sensitivity of the behavioral bias. Contingent on significance of one of the bias effect terms in the model, several *a priori* planned follow-ups were performed to better characterize the bias effect – a difference wave plot and an estimate of the bias effect at voiced, ambiguous, and unvoiced VOTs. To simplify the estimate of the bias effect by voicing ambiguity, the nine-VOTs are reduced into three ranges based on how they were perceived: short VOTs (5, 10, and 15) that for most listeners were judged as clearly voiced, mid-range VOTs (20, 25, 30) which were ambiguous, and long VOTs (35, 40, and 45), which were judged as clearly unvoiced.

Following the main experiment results two important follow-up analyses are presented. Section 4 evaluates a non-interactive alternative explanation for the N100 bias effects, whether N100 bias effects might be explained by perceptual learning. Section 5 evaluates how well the main experiment results match with the predictions of an interactive cognitive theory, TRACE (McClelland & Elman, 1986), as instantiated in the jTRACE model (Strauss, Magnuson, Harris, 2006).

3. Results

3.1 Behavior: Categorization Task Responses. Every subject showed a lexical bias effect in categorization, reporting the percept that formed a word more often than the percept that did not. The response data are presented in categorical perception curves in Figure 1 depicting the rate at which subjects perceived the unvoiced percept as a function of voice onset time for each continuum.

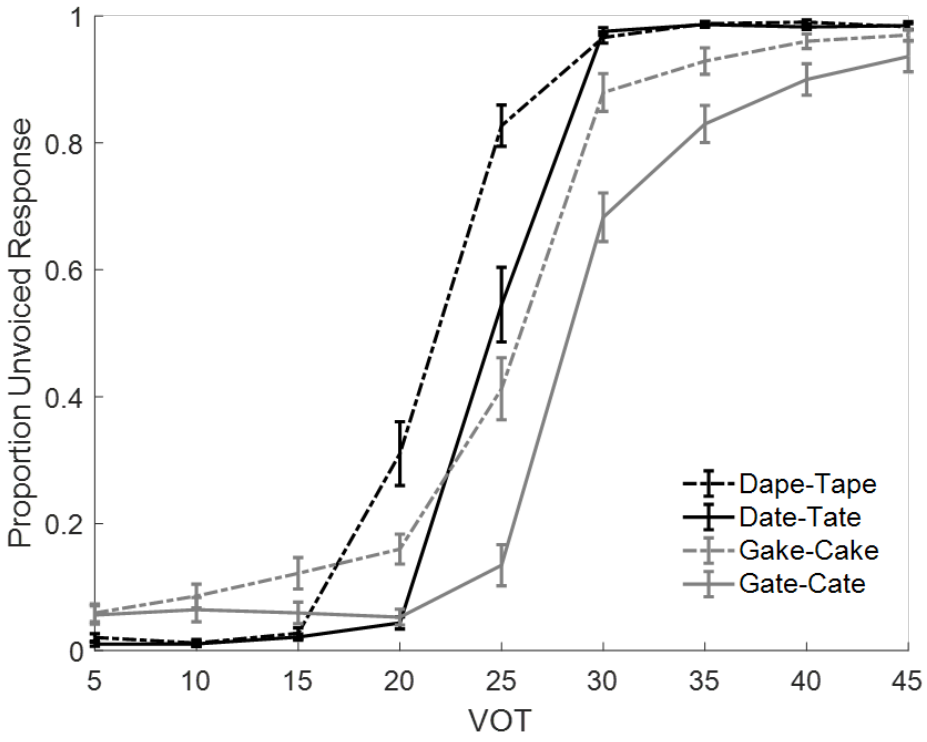

Figure 1. Results from categorical perception task. Lines depict the average proportion unvoiced response for each continuum at each VOT step. Unvoiced biasing continua are depicted with dashed lines. Voiced biasing continua are depicted with solid lines. Error bars depict the standard error of the mean. Note that the Ganong lexical bias effect is especially prominent at the mid-range VOTs 20, 25 and 30.

Table 1
Best Fitting Model for Subject Response Data

Fixed Effects	Beta	SE	t-statistic	95% CI
Intercept	-6.428***	0.367	-17.53	[-7.146, -5.709]
VOT	0.259***	0.003	97.69	[0.253, 0.264]
Bias	0.316***	0.036	8.86	[0.246, 0.386]
VOT-transform	-0.076	0.052	-1.49	[-0.178, 0.024]
Bias:VOT-transform	0.271***	0.053	5.17	[0.168, 0.374]
Random Intercepts:	N	Variance	SD	
Subject	20	0.172	0.415	
PlaceofArticulation	2	0.196	0.492	
TrialInBlock	72	0.014	0.120	

Notes. *N = 38,491 trials.* SE = Standard Error, SD = Standard Deviation. See Methods section for effect term codings. Responses were coded as 0 = voiced response, 1 = unvoiced response.
***p < .001

The best fitting model for the behavioral response data (responses coded as 0 = voiced response, 1 = unvoiced response) is reported in Table 1. In the data, effects of *VOT*, of *Bias*, and a *Bias:VOT-transform* interaction were evident. The positive estimate for the interaction of *Bias* with *VOT-transform*, i.e., stimulus ambiguity, indicates that the effects of lexical bias are largest for mid-range VOT stimuli. The main effect of *VOT-transform* was not significant but was included in the model since the interaction with *Bias* was significant. These results match the lexical bias effect in Ganong (1980), both that there is an effect of lexical bias on categorization in the predicted direction and that the effect of bias is greatest for the most ambiguous stimuli.

3.2 Omnibus Mixed Effect Model Results, N100. Before discussing the VOT and bias effect on the N100, we report the results of the forward stepwise model fitting procedure for N100 amplitude. This omnibus best fitting model is shown below in Table 2. In Section 3.3 and 3.4, we discuss the specific VOT and Bias effects of interest. Note that in the best fit N100 amplitude model reported above, *Bias* and *VOT-transform* main effects were maintained in the model despite not contributing significantly to the fit, in order to separately calculate the variance attributable to the interaction rather than having the interaction potentially convoluted by main effect variance.

Table 2
Best Fitting Model for N100 Amplitude

Fixed Effects:	*Beta*	*SE*	t-value	95% CI
Intercept	-0.79***	0.154	-5.15	[-1.09, -0.488]
VOT	0.0045**	0.0015	3.02	[0.0016, 0.0074]
Bias	-0.0018	0.0246	-0.08	[-0.050, 0.046]
VOT-transform	0.043	0.055	0.76	[-0.065, 0.152]
Bias:VOT-transform	0.124*	0.055	2.24	[0.016, 0.233]

Notes. *N = 38,491 trials. SE* = Standard Error. Variables were minimally recoded, see the Methods section for details. Random Effects for *Subject*, *Place of Articulation*, and *Trial In Block* were also included in the best fitting N100 model.
***p < .001, **p < .01, *p < .05

3.3 N100: VOT Encoding in N100. The N100 amplitude was sensitive to the VOT of the incoming sound. As shown in Figure 2, we observe a similar N100 - VOT relationship to that observed by Toscano et al. (2010). The mixed model indicated a significant linear relationship between stimulus VOT and N100 amplitude in the direction expected ($\beta_{VOT} = 0.0045, p = .003$). As reflected in the positive slope, speech sounds

with earlier onset of voicing, i.e., shorter VOTs, generated a larger N100 amplitude, i.e., more negative, than speech sounds with later onset of voicing. As discussed in depth in Toscano et al. (2010), this linear relationship indicates that the neural processes indexed in the N100 have not yet been warped by categorical aspects of perception such as phoneme classification.

The N100-VOT relationship evidenced by the significant *VOT* main effect shows that even in this paradigm with a slightly different sound stimulus set and task the N100 amplitude continues to show sensitivity to sublexical aspects of speech sound encoding. Further, the replication of the directionality of the VOT-N100 relationship yields directional hypotheses about the effects of bias. Since shorter voicing onset times corresponds to larger N100 amplitude, interactive theories predict that at the same VOT

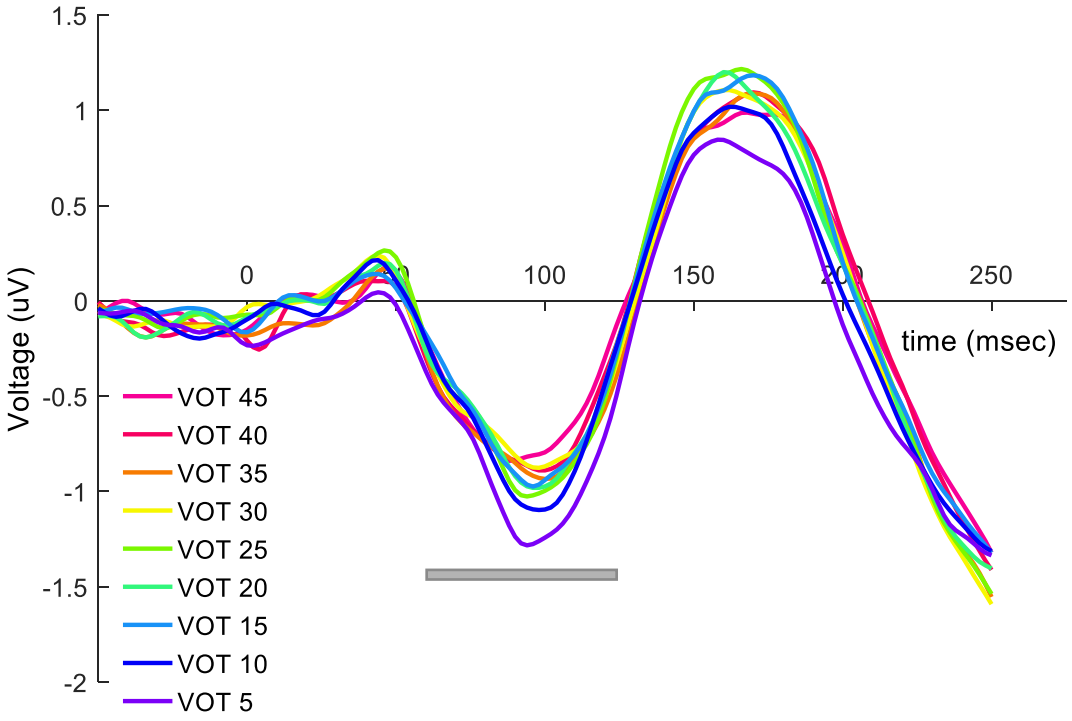

Figure 2. The ERP response to each VOT, averaging across all continua and Ganong Bias conditions is shown in each curve. The shorter the VOT, the more negative the N100 peak, as estimated by average amplitude from 60-130 msec. 0 ms = stimulus onset. Grey bar at bottom depicts the 60-130 msec time window.

level, continua that bias towards perception of a voiced consonant should have larger N100 amplitudes than continua that bias towards perception of an unvoiced consonant.

3.4 N100: Lexical Bias Effect.

Lexical bias effects were observed in N100 amplitude as shown in Figure 3. Lexical bias effects in the N100 were tested by evaluating the addition of a *Bias* term and a *Bias:VOT-transform* interaction term to the model. The significant *Bias:VOT-transform* ($\beta_{Bias:VOTtr} = 0.124,\ p = .025$) interaction term indicated that bias was observed in the direction specified *a priori* – more negative for voiced biasing than unvoiced biasing continua, and the bias effect interacted significantly with the transformed *VOT*. Recall that, in the interaction, VOT was transformed to correspond to the prediction of the largest bias effect at mid-range VOTs, just as it was in the behavioral analysis. Because of this transformation, the significant positive weight of the *Bias*:*VOT-transform* interaction suggests a bias in the correct direction and with the correct VOT specificity. The *Bias*:*VOT-transform* term did not interact with place of articulation, indicating a consistent effect across both places of articulation used in the experiment. The *Bias* main effect was non-significant reflecting that the majority of variance explained by the lexical bias was captured by the interaction term.

The N100 lexical bias matches the pattern in the observations of Ganong (1980) and the patterns in the behavioral data in this experiment (reported in Section 3.1) that lexical bias exerts an influence on sublexical judgments only at ambiguous VOTs. As we test in Section 5, this VOT ambiguity dependence is also predicted by interactive feedback models of speech perception.

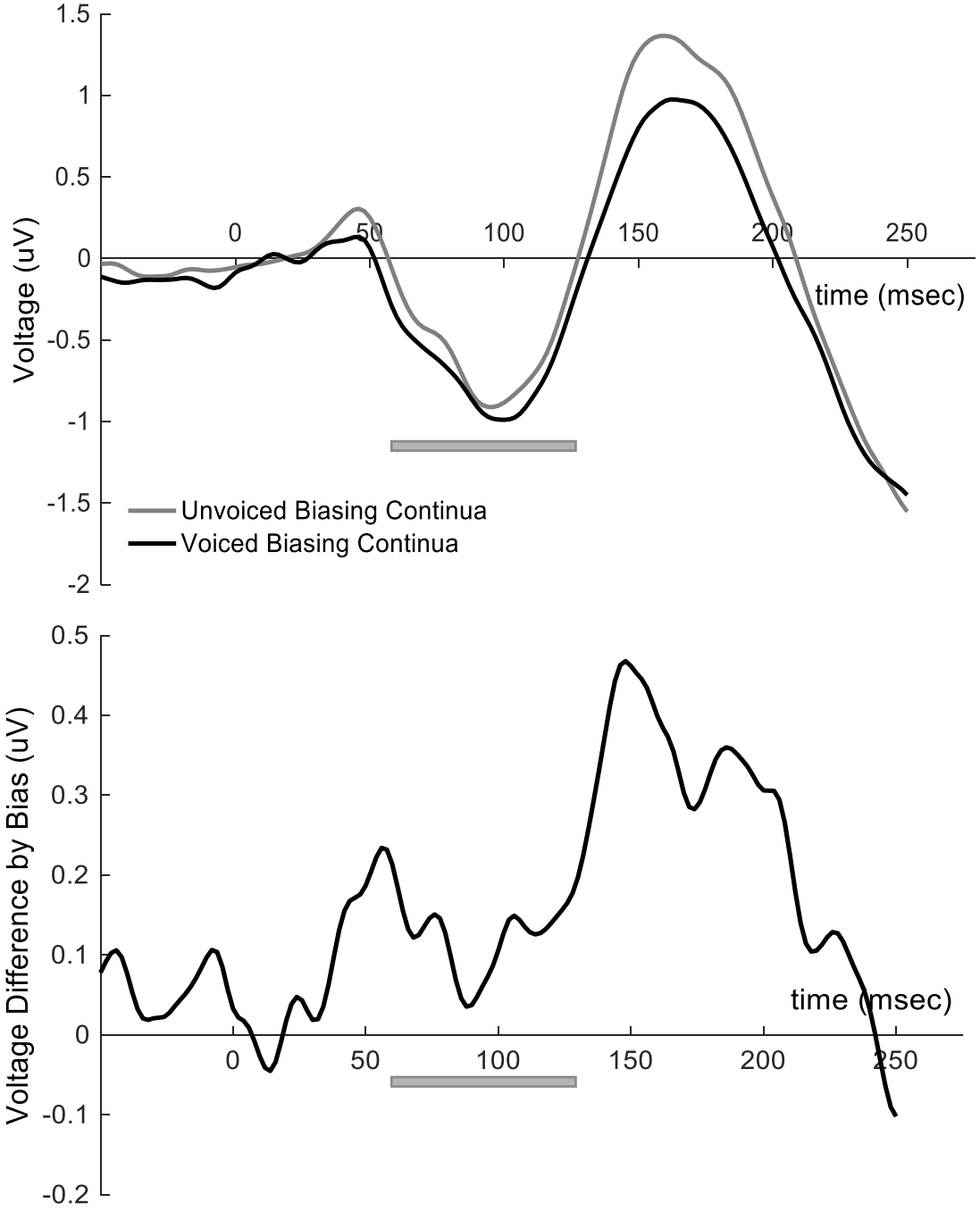

Figure 3. (top) N100 response amplitude for the voiced biasing and unvoiced biasing conditions to stimuli with ambiguous VOTs, 20, 25, and 30. Statistical analysis of bias effect focused on the 60 – 130 ms time window, depicted by the horizontal grey bar. (bottom) The difference wave estimate compares the voltage in the unvoiced and voiced biasing conditions in response to mid-range VOT stimuli.

3.5 N100: *Bias:VOT-transform* Interaction, Simple Effects Tests.

The positive slope of the *Bias:VOT-transform* interaction term in the omnibus mixed effects model indicated that the bias effect was largest at the mid-range VOTs. To better characterize this interaction, we wanted to estimate the size of the bias effect at the ambiguous VOTs as compared with the short VOTs (unambiguous voiced) and the long VOTs (unambiguous unvoiced). To obtain this estimate, mixed effect models were fit separately to the data from short VOT trials, mid-range VOT trials, and long VOT trials. In these restricted VOT ranges, the *Bias* main effect rather than the *Bias:VOT-transform* interaction is the term of interest, because we expect a roughly consistent

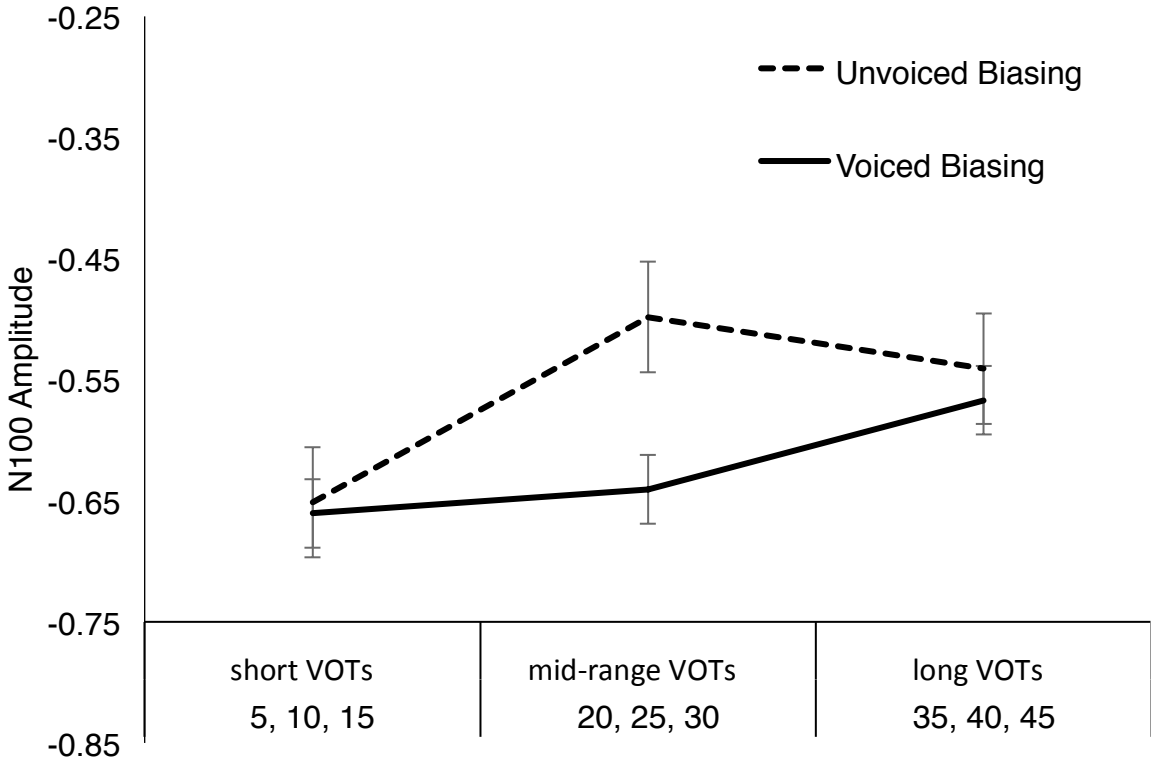

Figure 4. Average amplitude from 60-130 msec by VOT range and biasing direction. Error Bars depict Cousineau (2005) adjusted standard error of the mean. The bias effect is only significant for mid-range VOT trials.

bias effect size within each VOT range. The simple effect model was estimated with only the fixed effect of VOT and bias and the random effects from the best fitting model. This approach is analogous to simple effect tests following the interaction in an ANOVA. These simple effect models are presented as a follow-up to help interpret the results from the main analysis. As is visible in Figure 4, and as was suggested by the positive interaction in the omnibus model, the *Bias* effect was only significant at mid-range VOTs and no trace of a bias effect was present at short or long VOTs. The results for the *Bias* effect estimate at each VOT range are shown below in Table 3. The absence of the bias effect at the unambiguous VOTs is striking compared to the large bias effect at mid-range VOTs. This pattern makes clear why only the interaction term in the omnibus model reached significance.

Table 3
"Simple Effects" Model of Bias in Each VOT Range

VOT Range	$Beta_{Bias}$	*95% CI*	*N*	*Model Intercept*
short VOTs	0.0049	[-0.059, 0.069]	12,762	-0.727
mid-range VOTs	0.081*	[0.015, 0.15]	12,863	-0.642
long VOTs	0.012	[-0.053, 0.077]	12,866	-.0621

Notes. *N = number of trials in each VOT range. SE* = Standard Error. Simpe effects model included only fixed effect of *Bias* and random effects for *Subject, Place of Articulation*, and *Trial In Block.*
*p < .05

3.6 Interim Summary. In the N100 Ganong experiment reported in Section 3, we found effects of voice onset time and of lexical bias on the N100 waveform amplitude, providing evidence that as the sound is processed top-down information is interacting with incoming acoustic information to modulate the aggregate responses of the neural

populations encoding voicing. The N100 lexical bias effect provides a clear measure of when lexical bias effects occur in speech sound processing. The N100 lexical bias demonstrates, more clearly than behavior alone could, that the lexical bias effect is early, online, and sublexical rather than late and at a response selection stage. The top-down modulation of N100 amplitude is strongest when the bottom-up signal is ambiguous – i.e., the bias effect is sensitive to the ambiguity of the incoming voice onset time (see Figure 4 and Table 3). The VOT sensitivity of the bias effect, as well as the apparent growth of the bias effect size visible in the difference wave (Figure 3), indicates that bias is occurring during the trial and is not reflective of a preparatory activation of the favored sublexical units. Instead, it appears to reflect online feedback. Following the logic of Ganong (1980) though applying it here to neural data, the VOT sensitivity of the N100 bias effect confirms online versus preparatory activation because the information about the ambiguity of voicing of the incoming sound is not available until the sound itself is partially processed.

This pattern of results supports interactive theories of speech perception, which predict that sublexical processing should be influenced by top-down factors like lexicality and that this influence should be evident in online measures of sublexical processing such as the N100. These results contradict the predictions of feed-forward theories which have previously explained Ganong lexical bias effects in behavior by a post-perceptual merger of lexical and sublexical information (e.g., Fox, 1984; Norris, 1991; Norris et al., 2003, 2008, 2016) which therefore predict no lexical bias effect at this at this level of representation, at the level of phonetic feature encoding, or at this early time-point well before response selection. We view the body of results presented

here as favoring an interactive account of Ganong bias. However, as Norris and colleagues (2003, 2008, 2016) point out, results that are consistent with interactive theories may occur because of other mechanisms in a feed-forward architecture, for example because of perceptual learning effects that are occurring between trials. In Section 4 we test this perceptual learning account, to determine whether it can explain the results of lexical bias on the N100, without assumptions of interactivity.

4. Follow-Up 1: Evaluating Perceptual Learning as an Alternative Explanation for N100 Bias. In this section, we anticipate an important alternative explanation for N100 lexical biasing – perceptual learning. Perceptual learning is a mechanism of speech perception by which listeners adapt to and learn specific pronunciations for a speaker or listening environment. A perceptual learning explanation of lexical bias hypothesizes that by presenting participants with lexically biasing continua, participants may be learning or adjusting sublexical category boundaries to favor the lexically supported percept in each block. Perceptual learning is compatible with feed-forward models of speech perception, and accordingly must be ruled out for N100 bias effects to provide strongest evidence against feed-forward models of speech perception. At face value, a perceptual learning account may be particularly plausible given the design of our experiment. Trials were blocked by continua, meaning that over the course of approximately two minutes, participants hear 72 tokens all taken from one continuum. The ambiguous tokens in a dape-tape block, for example, are all lexically biased towards /t/, and therefore, over the course of these short blocks, participants might modify their phonemic categories by perceptual learning, such that the ambiguous sounds are remapped to be categorized as more /t/-like. If perceptual learning modifies

the same sub-lexical networks indexed by the N100, then perceptual learning is an alternative explanation for N100 bias effects. However, to anticipate the results, we observed no evidence for perceptual learning in the N100 lexical bias effect.

Critically, perceptual learning does not make identical predictions to online feedback; online feedback differs from perceptual learning in how the biasing accumulates with exposures (i.e., in perceptual learning it takes time and exposures to learn) and in the stability of the changes to the sublexical network (i.e., perceptually learned adjustments are stable across time and should generalize to other voicing continua). Contrastingly, in online feedback lexical activation is essentially instantaneous once the lexical target is known and can be similarly extinguished by reducing the activation of the lexical unit once a block is over.

We exploit these differences between perceptual learning and online feedback to test which is a better explanation for N100 bias in this experiment. Specifically, if the N100 bias effect we are observing in the experiment is due to perceptual learning, we expect that the size of the lexical bias should grow over the course of the 72 trials which comprise each block. Furthermore, since perceptual learning has been shown to be stable over short intervals and to generalize to similar acoustic contexts (Kraljic & Samuel, 2006, 2007; Bradlow & Bent, 2008), adjustments to voiced/unvoiced boundaries learned over one block should generalize to the next block. Because the experiment was built such that sometimes the previous block had the same bias and sometimes it had the opposing bias, a perceptual learning account would predict influences of the bias from the previous block on the current block lexical bias effect, at

least in the first portions of the following block. We empirically evaluated these two predictions of the perceptual learning account.

4.1 Approach. The significance of the learning build-up and carryover effects were tested within the mixed model by evaluating the addition of fixed effect terms reflecting each of these predictions. Specifically, we tested if the bias effect in the N100 amplitude model interacted with a term indexing the number of exposures to trials in the current block (learning build-up), and we tested if the previous block's bias could be detected in trials of the following block (carryover). For learning build-up, we tested if a model that included an interaction of bias with a count of exposures to trials in the current block improved model fit, relative to a base model, that did not contain the learning interaction[2]. A similar comparison approach was taken to test if a fixed effect term reflecting the previous block bias altered the N100 amplitude in the current block. If the N100 bias effect reflects perceptual learning, then growth of the bias effect by exposures in a block and carryover from the previous block should be evident in the N100 data. We also carried out a parallel analysis, looking at learning and carry-over in the behavioral responses.

4.2 Results. The full set of results of the perceptual learning model comparisons are reported in the Supplemental Material. A brief summary is provided here.

In the critical test of perceptual learning for explaining N100 bias, we found no evidence of perceptual learning from either prediction in the ERP data; that is there was

[2]To avoid the interaction capturing variance relating only to a main effect of trial number in block – i.e., how N100 amplitude changes over a block, systematically getting smaller in amplitude as neural responses do with repetitive inputs (e.g., Rabovsky, Hansen, & McClelland, 2018) – a main effect of *TrialNumberinBlock* was added alongside the interactions.

evidence for growth of the lexical bias with exposures to a block, nor an effect of the previous block bias carrying over. Perceptual learning terms decreased the fit of the model relative to models without these terms, and perceptual learning effect slope estimates never reached significance. The lack of fit of perceptual learning was evident even when non-linear learning rate functions were evaluated in the model (e.g., square root of exposure count). Even when we restricted our analyses to the first trials of each block where learning of the current block bias might be most evident and previous block carryover should be strongest, the predictions of perceptual learning were not supported in the data. Comparing the bias estimate from the first quarter of each block, on average just the first 6 ambiguous trials, the block bias effect estimate from these trials (i.e., the bias estimate from trials only in the first quarter of each block), is larger than the bias estimate for the trials which appeared later in the block (i.e., quartiles 2 - 4).

In the behavioral response data, we find a similar lack of support for perceptual learning. The perceptual learning effect terms never reached significance in the behavioral response model, and there was no evidence from the quartile models that the bias effect changed size with exposure count. While the behavioral response data are not as informative as the N100 data for whether N100 lexical bias is attributable to perceptual learning, they provide converging evidence that perceptual learning did not play a major role in shaping the behavioral responses, and therefore support the notion from the N100 perceptual learning tests that subjects' responses were better characterized as reflecting online feedback than perceptual learning.

4.3. Discussion of Perceptual Learning. Based on the failure of the perceptual learning predictions – learning with exposures and carryover – to fit the N100 lexical

bias data, we are unable to support perceptual learning as an alternative to online feedback for the N100 bias effect results of the main experiment. The pattern of lexical bias in the N100 data are much more compatible with the predictions of an interactive feedback account of Ganong bias than a perceptual learning account because an interactive account can easily accommodate an instantaneous activation of the current block's lexical bias and no carryover between blocks. Perceptual learning certainly plays a critical role in normal perception where we encounter multiple acoustic environments with multiple speakers with varying pronunciations (e.g., Bent & Bradlow, 2008). However, in the context of this N100 Ganong paradigm in which we have repeated exposure to the same speaker in multiple directions switching rapidly between blocks, perceptual learning seems to have not played a major role in determining subjects' neural and behavioral responses.

5. Follow-Up 2: TRACE simulation of ambiguous VOT locus of Ganong Bias

In this section, we ask how well the N100 bias effects in this experiment match lexical bias effects predicted in an interactive theory of speech perception, TRACE (McClelland & Elman, 1986), as instantiated in the jTRACE model (Strauss, Harris & Magnuson, 2007). Specifically, we examine whether such a theory predicts that the effect of lexical bias on a sublexical level is greatest for the most ambiguous sounds.

5.1 Modeling Methods and Stimuli. The standard implementation of jTRACE (Strauss et al., 2007) was accessed via GitHub, and was run using the default lexicon (BigLex) and standard parameter settings. We selected /d/-initial and /t/-initial words which formed word-nonword and nonword-word pairs as defined by which words exist in the TRACE lexicon and which did not. The final modelling continua selected were Dal-Tal

(/d/-baising since Doll is a word in the TRACE lexicon and Tall is not) and Dar-Tar (/t/-biasing since Tar is a word in the TRACE lexicon and Dar is not). They were chosen over other pairs of this type because they have similarly dense cohorts on each side of the d-t continuum and because the coda position phoneme in both is a liquid, to minimize accidental activation of the /d/ or /t/ units by the coda phoneme. /d/-/t/ voicing continuum inputs to the model were prepared using the ambiguous phoneme tool in jTRACE to create inputs that model the VOT manipulation of the ERP experiment. Activation of the /d/ and /t/ phoneme units in response to these inputs were measured from processing steps 1-75, with activation values time-locked to the specified input step.

5.2 Analysis Methods for Modeling Results. Lexical bias effects were examined in the phoneme layer of the TRACE model. The phoneme layer is the best candidate to model the sublexical level indexed by the N100. Although the feature layer of TRACE might more closely correspond to the feature encoding evident in the N100, in TRACE there are not feedback connections down to that level, so feedback cannot be modeled at that level of the model. Within the phoneme layer, we quantified lexical bias as the difference in activation of the /t/ and /d/ units in response to /d/-/t/ inputs embedded in the lexically biasing pairs. Since lexical bias in our experiment is defined by comparison with a lexical environment of opposite bias, a similar estimation of lexical bias was calculated for the model. That is, lexical bias was estimated by comparing the /d/ vs. /t/ activation to each lexical bias direction. Thus, model bias was calculated in a two-step procedure for each VOT step, *i*, on the /d/-/t/ continua:

$$\mathrm{Bias}_{\mathrm{step\ i}} = (/\mathrm{t}/-/\mathrm{d}/\ \mathrm{activation}_{\mathrm{dar-tar}_{\mathrm{step\ i}}}) - (/\mathrm{t}/-/\mathrm{d}/\ \mathrm{activation}_{\mathrm{dal-tal}_{\mathrm{step\ i}}}) \quad (1).$$

First, the /t/-/d/ phoneme activation difference was calculated for each VOT step, *i*, for each bias direction continuum. In the second step, the /t/-/d/ activation difference for dal-tal, /d/ biasing, was subtracted from dar-tar, /t/ biasing. A positive value for the bias effect indicates that the /t/-/d/ activation difference was larger when the lexicon was /t/-biasing than when the lexicon was /d/-biasing. Phoneme activations were calculated using the specified alignment method in jTRACE.

5.3 Modeling Results.

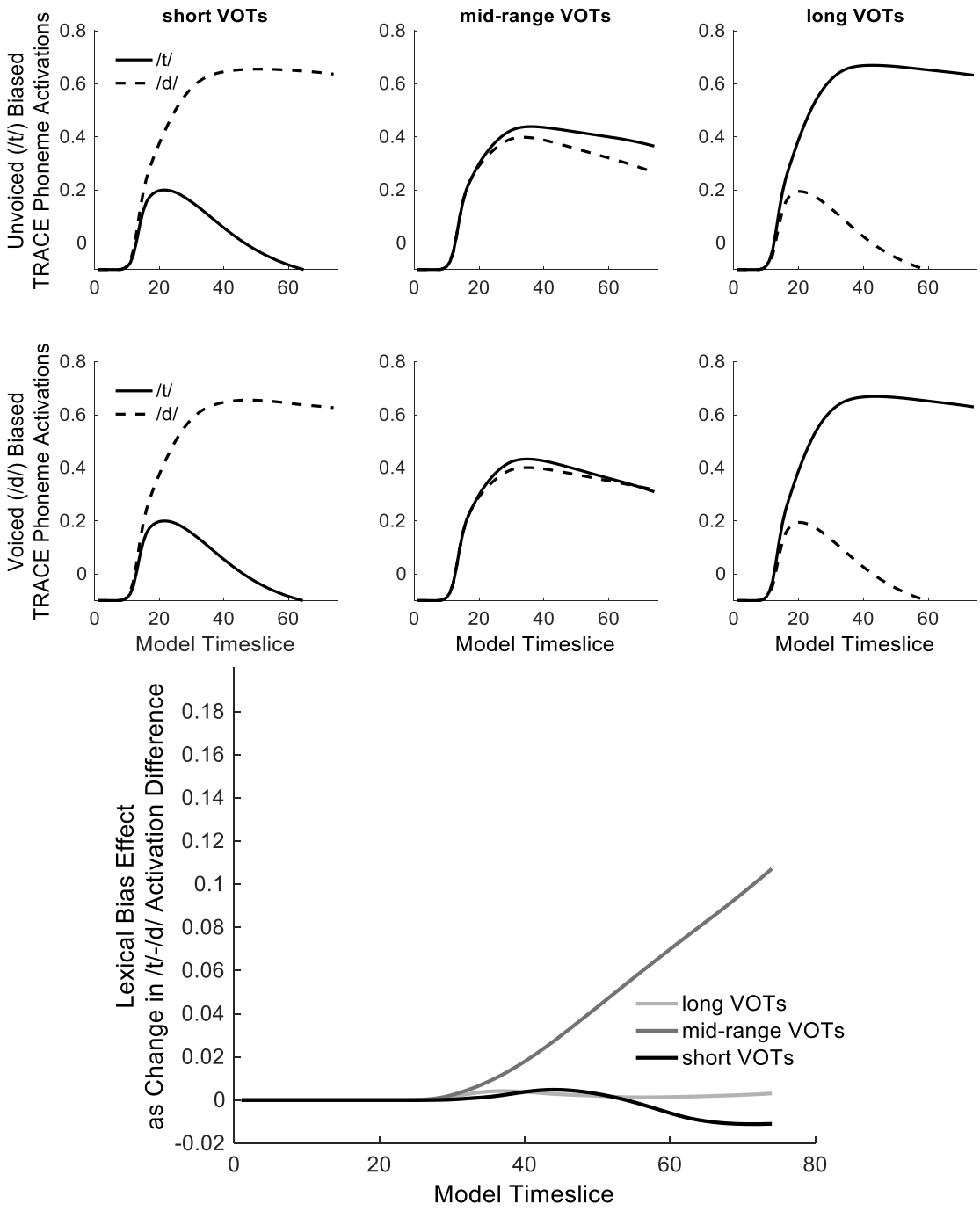

Figure 5. (top) Model phoneme activations for each input ambiguity range and biasing condition. (bottom) Lexical Bias effect, as defined in Equation 1, comparing d and t activation difference by biasing condition. Note that the bias effect is only evident in the ambiguous input to the model. The model lexical bias effect emerges around timeslice 30, and then continues to grow as processing moves forward. Specified alignment activations in jTRACE were used. Note that TRACE does not allow manipulation of VOT or sound file inputs; thus, VOT equivalent model inputs were created using the ambiguous phoneme tool in jTRACE.

A lexical bias on phoneme level activations was observed. The bias on phoneme activations was in the direction expected. As a more fine-grained test of whether TRACE can model N100 lexical bias effects, we found that the lexical bias effects in TRACE match the patterns in the N100 data – namely that lexical bias develops during processing and that lexical bias is largest when the input is ambiguous. Figure 5, bottom, demonstrates both phenomena in that the lexical bias effect grows as model processing moves forward in time, and that the model lexical bias effect is only present in a significant manner at ambiguous inputs. This pattern of lexical bias only at ambiguous model inputs exactly matches the results obtained in the ambiguous VOT locus of the N100 bias. The growth of the lexical bias effect in the model as processing progresses also matches that the N100 lexical bias effect grew rapidly from 75-175 msec. While it is difficult to exactly map time-slices in the model onto neural processing times, it suggests compatibility of the N100 bias with the dynamics of how feedback alters lower level representations in interactive models.

5.4 Modeling Summary. The modeling results demonstrate how an interactive speech perception framework (TRACE) can model Ganong bias at the phoneme level. The model results closely match several aspects of the N100 bias results. First, the largest bias effect in both the model and in behavior was in the case of ambiguous inputs and lexical bias was essentially nonexistent if the model input was unambiguous. Second, the bias effect within the sublexical units grows over the course of processing.

General Discussion

In the current study, we collected electrophysiological responses to voice onset time continua while participants took part in a lexically biasing categorical perception

experiment. We discovered an effect of lexical bias within the N100 ERP component, a signal associated with encoding of the phonetic feature of voicing. Lexical bias on the N100 amplitude and on categorization was strongest only when the incoming sounds were ambiguous, near the boundary between voiced and unvoiced, matching the predictions of interactive theories. The sensitivity of bias effects to the ambiguity of the bottom-up acoustic information suggest that the bias effect is online, rather than reflecting pre-activation of the lexically favored sublexical unit. This follows from the fact that ambiguity of the incoming stimulus is not available until processing of the acoustic information is partially completed.

Strengthening the claim that the N100 bias reflects a lexical bias effect within the sublexical network is the fact that the biased N100 responses match the directionality of the normal responses to voiced and unvoiced endpoints. For example, in a /t/-biased continuum, an ambiguous /dt/ was shifted towards the response to a normal /t/. That the /t/-biased response to an ambiguous phoneme resembles the response to a normal /t/ phoneme suggests that lexical biasing of perception is accomplished by activation changes within the normal sublexical processing regions. The same was true for /d/ biased ambiguous stimuli.

Further supporting interactive feedback accounts of lexical bias, two attributes of the empirical N100 data – that the bias effect is limited to ambiguous tokens and that the bias effect grows during the course of sublexical processing– match the predictions of an interactive feedback model of speech perception (Elman & McClelland, 1988). These aspects of the N100 data were also fit well in our own modelling experimentation, reported in Section 5. Using the standard parameters and lexicon (Strauss, Magnuson,

& Harris, 2007) and taking phoneme activation as a proxy for N100 level processing, both aspects of the N100 lexical bias emerged naturally from TRACE (McClelland & Elman, 1986) when inputting a word-initial, acoustically ambiguous (in model voicing feature-space) lexically biased continuum.

The N100 lexical bias results are difficult to reconcile with previous feed-forward accounts of lexical bias in behavior. We find clear evidence against the established feed-forward explanation for Ganong lexical bias effects, that lexical bias reflects lexical influence at a post-perceptual response selection stage in which lexical knowledge and bottom-up information are merged late in speech sound processing. This post-perceptual account of Ganong lexical bias is incompatible with at least three aspects of the N100. First, the time-course of lexical bias effects is early (approximately 75-175ms). Previous research has suggested that this is the time course of processing acoustic and phonological information (Cibelli et al., 2015; Hullet et al., 2016; Yi, Leonard, & Chang 2019; Mesgarani et al., 2014; Pasley et al., 2012). Second, lexical bias effects appear with the same topography and polarity as normal N100 responses to speech sounds, suggesting that it is the same neural populations that show the lexical bias that are also responsible for the basic representation of the speech sound phonetic features (see Myers & Blumstein, 2007, for a similar argument using fMRI). As a result, it is unlikely that the N100 is indexing activation of the lexicon directly, but instead is revealing the influence of lexical activation on the sublexical level. Third, the bias effect grows during processing, which are the dynamics predicted by interactive theories but not feed-forward accounts with a post-perceptual merger.

All this is to say, the N100 evidence is inconsistent with a post-perceptual explanation of Ganong lexical bias. The N100 bias effect demonstrates that lexical bias effects are due to changes in the early processing of speech sounds, specifically in changes to the very same networks involved in encoding phonetic features of the unbiased phonemes. This result can be viewed as evidence against the feed-forward principle at the core of the post-perceptual hypothesis, that early sublexical encoding is isolated from top-down information. This independence of early processing principle is contradicted by the neural response data from the current study and that of Getz & Toscano (2019). Instead, there is every indication that lexical knowledge is influencing the earliest measurable aspects of sublexical encoding in exactly the ways predicted by interactive models of speech perception with feedback.

Of course, there are some ways that feed-forward theories allow for sublexical encoding to be modified by the lexicon, specifically by perceptual learning. As such, in Section 4, we also evaluated perceptual learning as an alternative mechanism for Ganong lexical bias. Under this account, listeners are constantly and dynamically changing their representations of speech sound categories, based on numerous factors including information about lexicality (Norris, McQueen & Cutler, 2003). However, in a series of mixed effects models, we failed to find support for any of the predictions of a perceptual learning account for the lexical bias effects in this experiment, demonstrating again that the results of our study are inconsistent with feed-forward theories of speech perception, even when multiple possible feed-forward explanations are considered.

Therefore, the theoretical contribution of this study is clear. By looking at an early electrophysiological correlate of sublexical speech perception, we have been able

to demonstrate that lexical context exerts a direct influence on sublexical processing, as would be predicted by interactive models of speech perception, in direct contrast to feed-forward only models of speech perception. This study provides clear evidence sublexical encoding is not independent as feed-forward theories predict. Feed-forward theories have previously argued that interactivity of lexical knowledge with bottom-up acoustic information may result in loss of veridical information about the true acoustics of the input, and that optimal Bayesian use of information, from a theoretical information analysis perspective, is achieved by a feed-forward architecture (Norris et al., 2016, though see McClelland, Mirman, Luthra, Strauss, & Harris, 2018 for a counterargument). Norris and colleagues (2000, 2003, 2008, 2016) have also argued that feed-forward architectures are simpler than interactive ones. However, optimality and debates of theoretical simplicity must yield to empirical evidence. At least on the scale of neural activity that we can index with ERP N100 amplitudes, we found modulation of the sublexical level of representation by lexical information. The N100 evidence is clear that there are at least some neural populations, with dipoles that align with VOT encoding populations, that are responding to top-down information. The claim that early, sublexical speech sound processing is independent from contextual information is not compatible with this evidence.

This study is not alone in using electrophysiology to investigate sublexical contextual effects. There are several other recently published studies that have measured the electrophysiological signature of sublexical processing and shown evidence of feedback from contextual information (e.g., Getz & Toscano, 2019; Leonard, Baud, Sjerps, & Chang, 2016). These interactions follow exactly the patterns

predicted by interactive theories of speech perception (McClelland & Elman, 1986), with greater activations observed for the contextually favored sublexical component. While there have been several recent studies that have demonstrated similar effects to what we report here, it is worth noting how our paradigm and results complement, rather than simply replicate, these other studies. Leonard and colleagues (2016) played participants pairs of spoken words that were acoustically identical except for a critical phoneme that differentiated their meaning (e.g., “factor” vs. “faster”), as well as an ambiguous token that replaced that critical phoneme with broadband noise. Approximately 100 msec following the onset of the ambiguous speech sound, the pattern of ECoG data in bilateral auditory cortex predicted which of the two possible words the participant would report hearing, identifying the time course of phoneme restoration effects. Getz and Toscano (2019), instead, relied on lexical prediction, looking at the response at the N100 to an ambiguous speech sound between /d/ and /t/ in cases in which a /t/ would be predicted (“Eiffel Tower”) and cases in which a /d/ would be predicted (“Barbie Doll”). They found that the N100 amplitude to this ambiguous sound was significantly more /t/-like when primed with “Eiffel” than when primed with “Barbie”.

These three studies look at different types of contextual effects, which may tell us different things about the nature of sublexical processing. In the Leonard et al. (2016) study, lexical information is restoring missing acoustic information in the neural response patterns 100msec after the missing phoneme. In the current study and Getz & Toscano (2019), contextual information modulates the behavioral responses and show similar effects in N100 amplitudes, shifting perception to the contextually favored

state. The difference between our study and the Getz and Toscano (2019) study can be conceived of in terms of short-term vs. long-term contextual effects. In Getz and Toscano (2019), the reason that participants are biased to hear an ambiguous /dt/ as /t/ is because of a specific prediction about which upcoming word is expected, based on long-term associations between words within the lexicon. In the current study, lexical expectations were created online through the use of the blocked Ganong design, but these context effects were generated only over the short-term and would change between blocks. The lexical effects in this experiment also occur automatically, and do not involve prediction, *per se.* These different types of contextual effects may occur because of different computational properties of the speech perception system and therefore it is not obvious that each of these manipulations would lead to similar electrophysiological effects. This growing literature supports the idea that multiple types of contextual effects have an impact on the sublexical processes being indexed by the N100.

One direction for future research, therefore, is to use the N100 to determine if all types of contextual effects occur with the same time-course and at the same processing level. Behaviorally, there are many ways that context has been shown to influence our perception of speech sounds, whether it be at a lexical (Ganong, 1980), semantic (Samuel, 1981; Connine & Clifton, 1987; Groppe, Choi, Huang, Schilz, Topkins, Urbach, & Kutas, 2010), syntactic (Fox & Blumstein, 2016), or through indexical information, such as gender (Johnson, Strand, & D'Imperio, 1996) or accent (Bent & Bradlow, 2008; Sidaras, Alexander, & Nygaard, 2009). By using electrophysiology, we can measure the time-course and processing stage at which contextual effects are occurring, thus

gaining greater traction on questions about interactivity. Our study shows that at least one of these top-down effects, lexical biasing via the Ganong paradigm, is best understood by an interactive theory of speech perception. Getz & Toscano (2019) test another top-down effect and show a similar top-down effect on the N100 response. But for the myriad other top-down effects, electrophysiological work will be necessary to characterize the time-course and neural signature of each effect.

Of course, our ability to interpret what top-down effects in the N100 mean in terms of cognitive architectures depends deeply on our confidence in the relationship between the neuroimaged variable used in the study with the cognitive processing step it measures. This problem is common to most cognitive neuroscientific studies. With respect to the N100, it is clear that the N100 is indexing some aspect of sublexical encoding associated with the encoding of voice onset time. However, sublexical encoding is a multi-step process that may involve a gradual and parallel activation of spectral, spectrotemporal, featural, and phoneme representations. Here we interpret the N100 as indexing pre-categorical featural processing, as is suggested by Toscano et al. (2010) and electrocorticographic studies (Hullet et al., 2012; Mesgarani et al., 2014; Leonard et al., 2016) One alternative view could be that the N100 indexes acoustic (not language specific) levels of spectrotemporal processing that feed into our ability to recognize speech sounds but which precede linguistic processing. Effects of voice onset time on amplitude might be expected at an acoustic level of processing as well, because voicing is both a low-level acoustic feature as well as a linguistic property. Similarly, the N100 may reflect acoustic or sublexical linguistic processing. But if the N100 is prelinguistic, effects of lexical bias are even more surprising because this is an

earlier stage of processing. Therefore, this would make N100 bias even more contradictory with the feed-forward view that initial stages of bottom-up processing are independent from top-down influences.

Testing the precise cognitive correlate of the N100 is important for interpreting what the results of this and similar studies mean for cognitive architectures. Indeed, the claims made above depend on the assumption that the processes being indexed by N100 are sublexical and therefore not compatible with post-perceptual merger accounts of lexical bias. But it may not be so straightforward to map between specific time windows from the EEG signal and specific cognitive processes because in an interactive theory like TRACE, activation spreads from one level to another well before the computations at the first level are complete. As a result, all levels of representation are activated in parallel, meaning that there is processing of information simultaneously at acoustic, sublexical, and lexical and semantic levels.

That all levels are activated in parallel does not mean they cannot be distinguished electrophysiologically. The fact that the N100 response to voice onset time is similar in a wide array of lexical settings (e.g., the words used in this study, differ from those of Toscano et al., 2010 and from Getz & Toscano, 2019), suggests that the N100 does not reflect activation of a particular word, but rather a common acoustic or featural aspect of the speech sound. This distinction is important because it rules out the possibility that N100 lexical bias might simply reflect parallel lexical activation. Similarly, distinguishing between acoustic and featural levels and the contribution of each to the N100 will inform how evidence from N100 ERP components constrain theory.

One other area for further investigation involves increasing spatiotemporal resolution of the lexical bias effect. In our data, and in the modelling results, there is the suggestion that lexical bias grows as processing moves forward in time (from 75-175 msec after the generating phoneme). An intriguing way to interpret this data would be to say that the informational representations in a single network or a single population of neurons evolves over time, specifically that the neural response of the N100 generating region is initially determined by the bottom-up input in response to voice onset time but gradually shows a greater response to top-down information. To be able to evaluate this claim more precisely, one would need to test with greater spatial resolution that the same populations are changing which sources information they are responding to or how they are weighting multiple inputs over time (see Hirshorn and colleagues, 2016 for a demonstration of a similar phenomenon in the reading system). Techniques with better spatial resolution such as ECoG may be able to answer these questions by using the same experimental design to test if specific regions become more sensitive to top-down contextual information as the trial progresses.

To conclude, the benefits that the EEG/ERP experiments provide to the study of early perceptual processing are clear. They provide high temporal resolution measures that can track cognitive processes as they unfold over time, providing a new window into resolving old questions about when top-down and bottom-up information are integrated. Using EEG/ERP, specifically by analyzing the N100 response to voice onset time, we found that lexical information influences how we are processing speech sounds as early as 100ms after the onset of the stimulus. That an early neural response is influenced by lexical bias contradicts the idea that early processing of

speech sounds is independent from higher-level knowledge. Instead, we find support for interactive theories of speech perception, and interactive theories of neural and cognitive processing more generally.

Author contribution statement

CN and SF-B developed the study concept and study design. This study formed a portion of the master's degree project for CN. The study design and analysis plans were reviewed and approved by the master's degree committee prior to data collection. Testing and data collection and data analysis were carried out by CN under the supervision of SF-B. CN drafted the manuscript and SF-B provided critical revisions. All authors approve of the final manuscript for submission.

Open Science Practice Statement

Materials, software, behavioral and EEG data are published on PsyArxiv, hosted by OSF (at https://osf.io/8f7kj/ or by searching DOI: 10.17605/OSF.IO/8F7KJ). The data are in a matlab form (.mat), and a readme file in the same location describes the data labelling.

Acknowledgements

This project was supported by the National Science Foundation under Grant No. 1250104 to C.N. and Grant No. 1752751 to S.F-B. Jake Johnston, Georgia Jenkins, Ruthie Seleznick, Serena Brandler, and Talia Liu assisted in stimulus preparation and data collection. We would like to thank James Magnuson for his help in developing the TRACE simulations. We would also like to thank Bob McMurray and Joe Toscano for help with stimulus splicing techniques and study conceptualization.

Works Cited

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. Cognition, 52(3), 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*(3), 255-278.

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

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. Journal of Experimental Psychology: Human Perception and Performance, 15(3), 567.

Burton, M. W., & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. Journal of Experimental Psychology: Human Perception and Performance, 21(5), 1230.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. P. (1995). Bottom-up connectionist modelling of speech. Connectionist models of memory and language, 289-310.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. Cognitive Psychology, 33(2), 111-153.

Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. Cognition, 34(1), 93-107.

Cibelli, E. S., Leonard, M. K., Johnson, K., & Chang, E. F. (2015). The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. Brain and language, 147, 66-75.

Connine, C. M., Clifton Jr, C., & Cutler, A. (1987). Effects of lexical stress on phonetic categorization. Phonetica, 44(3), 133-146.

Connine, C. M., & Clifton Jr, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human perception and performance*, *13*(2), 291.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in quantitative methods for psychology*, *1*(1), 42-45.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory phonology*, *10*, 91-111.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. Journal of Cognitive Neuroscience, 23(12), 3914-3932.

Davis, C., Kislyuk, D., Kim, J., & Sams, M. (2008). The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. Brain research, 1242, 151-161.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950-1953.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. Journal of Memory and Language, 27(2), 143-165.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception and performance*, *10*(4), 526.

Fox, N. P., & Blumstein, S. E. (2016). Top-down effects of syntactic sentential context on phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 730.

Frauenfelder, U. H., & Peeters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance*, *6*(1), 110.

Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. Psychological science, 0956797619841813.

Gow Jr, D. W., Segawa, J. A., Ahlfors, S. P., & Lin, F. H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. Neuroimage, 43(3), 614-623.

Gow Jr, D. W., & Olson, B. B. (2015). Lexical mediation of phonotactic frequency effects on spoken word recognition: A Granger causality analysis of MRI-constrained MEG/EEG data. Journal of memory and language, 82, 41-55.

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain research*, *1361*, 54-66.

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. Frontiers in psychology, 4, 563.

Hirshorn, E. A., Li, Y., Ward, M. J., Richardson, R. M., Fiez, J. A., & Ghuman, A. S. (2016). Decoding and disrupting left midfusiform gyrus activity during word reading. Proceedings of the National Academy of Sciences, 113(29), 8162-8167.

Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., & Deltenre, P. (2009). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in French-speaking subjects. Clinical Neurophysiology, 120(5), 897-903.

Horev, N., Most, T., & Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non-speech (FOT) signals: behavioral and electrophysiological correlates. Ear and hearing, 28(1), 111-128.

Hornickel, J., Skoe, E., & Kraus, N. (2009). Subcortical laterality of speech encoding. Audiology and Neurotology, 14(3), 198-207.

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *Journal of Neuroscience*, *36*(6), 2014-2026.

Hutchison, E. R., Blumstein, S. E., & Myers, E. B. (2008). An event-related fMRI investigation of voice-onset time discrimination. Neuroimage, 40(1), 342-352.

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

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. Journal of experimental psychology: Human perception and performance, 42(12), 1969.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148.

Kokinous, J., Tavano, A., Kotz, S. A., & Schröger, E. (2017). Perceptual integration of faces and voices depends on the interaction of emotional content and spatial frequency. Biological psychology, 123, 155-165.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. Psychonomic bulletin & review, 13(2), 262-268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. Journal of Memory and Language, 56(1), 1-15.

Leonard, M. K., Baud, M. O., Sjerps, M. J., & Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature communications*, *7*, 13619.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology, 54(1), 146-157.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron, 54(6), 1001-1010.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018a). Interaction in spoken word recognition models: Feedback helps. Frontiers in psychology, 9, 369.

Magnuson, J., You, H., Nam, H., Allopenna, P., Brown, K., Escabi, M., ... & Rueckl, J. (2018b). EARSHOT: A minimal neural network model of incremental human speech recognition.

Martin, B. A., & Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. Ear and hearing, 20(1), 33-44.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception?. *Trends in cognitive sciences*, *10*(8), 363-369.

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. Cognitive science, 38(6), 1139-1189.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. Journal of Experimental Psychology: Human Perception and Performance, 17(2), 433.

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(7), 860-863.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes?. Journal of Memory and Language, 61(1), 1-18.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006-1010.

Myers, E. B., & Blumstein, S. E. (2007). The neural bases of the lexical effect: an fMRI investigation. *Cerebral Cortex*, *18*(2), 278-288.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. Journal of experimental psychology: Human perception and performance, 23(3), 873.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299-325.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review*, *115*(2), 357.

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. Language, cognition and neuroscience, 31(1), 4-18.

Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. Journal of Cognitive Neuroscience, 16(1), 31-39.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. Neuroimage, 20(3), 1839-1847.

Obleser, J., Rockstroh, B., & Eulitz, C. (2004). Gender differences in hemispheric asymmetry of syllable processing: Left-lateralized magnetic N100 varies with syllable categorization in females. Psychophysiology, 41(5), 783-788.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. Psychological review, 85(3), 172.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. PLoS biology, 10(1), e1001251.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. Psychophysiology, 52(11), 1456-1469.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4), 1037.

Pitt, Mark A., and Arthur G. Samuel. "Lexical and sublexical feedback in auditory word recognition." Cognitive psychology 29.2 (1995): 149-188.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. Journal of Experimental Psychology: Human Perception and Performance, 19(4), 699.

Pylyshyn, Z. W. (1984). Computation and cognition (p. 41). Cambridge, MA: MIT press.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. Nature Human Behaviour, 2(9), 693.

Rosburg, T., Boutros, N. N., & Ford, J. M. (2008). Reduced auditory evoked potential component N100 in schizophrenia—a critical review. *Psychiatry research*, *161*(3), 259-274.

Rumelhart, D. E. (1976). Toward an interactive model of reading. San Diego, California: Center for Human Information Processing, University of California.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. Cognitive Brain Research, 15(3), 228-240.

Schreiber, K. E. (2017). *Who Cares Who's Talking? The Influence of Talker Gender on How Listeners Hear Speech* (Doctoral dissertation, The University of Iowa).

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. The Journal of the Acoustical Society of America, 106(2), 1078-1083.

Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. The Journal of the Acoustical Society of America, 125(5), 3306-3316.

Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain research*, *1121*(1), 177-189.

Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special?. *Hearing research*, *305*, 57-73.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1994). Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Clinical Neurophysiology*, *92*(1), 30-43.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39,* 19-30.

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.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. Neuron, 102(6), 1096-1110.

Zaehle, T., Jancke, L., & Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. Behavioral and Brain Functions, 3(1), 63.

Appendix A

Stimulus Properties

All stimuli were embedded within a 600 msec .wav sound file. 100 msec of silence preceded the plosive burst. Across each opposing pair of continua, Gate-Cate/Gake-Cake and Date-Tate/Dape-Tape, endpoints were carefully matched.

Table A1. *Stimulus Acoustic Properties*

Continuum	Burst – Closure of Vowel (ms)	Burst – Release offset (ms)	Pitch (Hz)	CV Ratio	Frequency of Lexical Endpoint
Gate-Cate	210	430	220	.12	3
Gake-Cake	212	400	221	.12	2
Date-Tate	180	397	190	.14	11
Dape-Tape	186	376	212	.13	11

Note. Pitch was averaged across the vowel length; CV Ratio was averaged across all VOTs, effectively the CV ratio for VOT-25; Brown Verbal Frequency was used as frequency estimate.

Appendix B

Stimuli Pilot Study

One critical test necessary before running the full electrophysiological study was ensuring that the stimuli yielded clean categorical responses on both the voiced and unvoiced endpoints and ensuring that they yielded lexical bias effects on categorization. The stimuli were piloted in a separate set of subjects (N = 5). The results of this pilot study are plotted below in categorical perception curves. As is visible, categorization was good and lexical bias effects were evident.

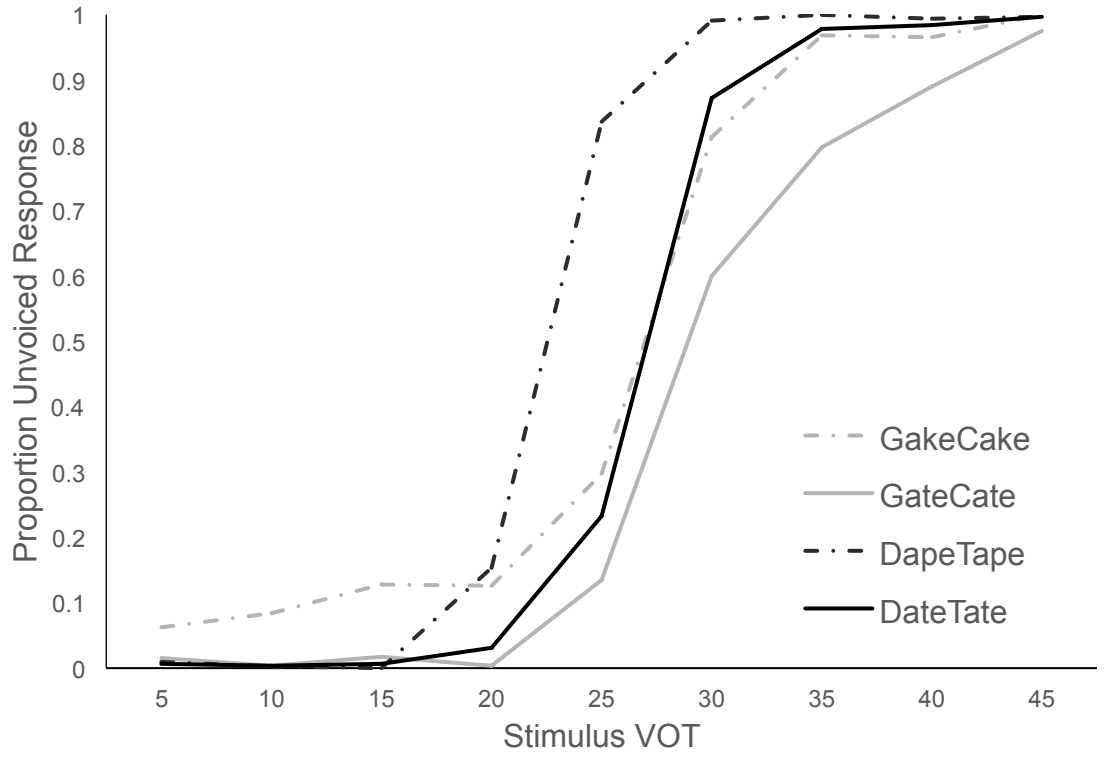

Figure A1. Proportion Unvoiced Response to each stimulus in a behavioral pilot experiment (N = 5). Subject pool was separate from the eventual full electrophysiological experiment.