Journal of Experimental Psychology:
Human Perception and Performance
2016, Vol. 42, No. 12, 1969–1988

0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000269

Eye Movement Evidence for an Immediate Ganong Effect

John Kingston, Joshua Levy, Amanda Rysling, and Adrian Staub
University of Massachusetts, Amherst

Listeners tend to categorize an ambiguous speech sound so that it forms a word with its context (Ganong, 1980). This effect could reflect feedback from the lexicon to phonemic activation (McClelland & Elman, 1986), or the operation of a task-specific phonemic decision system (Norris, McQueen, & Cutler, 2000). Because the former account involves feedback between lexical and phonemic levels, it predicts that the lexicon's influence on phonemic decisions should be delayed and should gradually increase in strength. Previous response time experiments have not delivered a clear verdict as to whether this is the case, however. In 2 experiments, listeners' eye movements were tracked as they categorized phonemes using visually displayed response options. Lexically relevant information in the signal, the timing of which was confirmed by separate gating experiments, immediately increased eye movements toward the lexically supported response. This effect on eye movements then diminished over the course of the trial rather than continuing to increase. These results challenge the lexical feedback account. The present work also introduces a novel method for analyzing data from 'visual-world' type tasks, designed to assess when an experimental manipulation influences the probability of an eye movement toward the target.

Keywords: speech perception, phoneme categorization, eye movements, Ganong effect

Speech sounds nearly always occur in the context of other speech sounds. The context in which a sound occurs may influence how the sound is perceived by activating linguistic knowledge of words (e.g., Ganong, 1980; Pitt & Samuel, 1993), transitional probabilities among sounds (e.g., Pitt & McQueen, 1998; Magnuson, McMurray, Tanenhaus, & Aslin, 2003), and probabilistic and categorical phonotactics (e.g., Massaro & Cohen, 1983; Vitevitch & Luce, 1998, 1999; Hallé, Segui, Frauenfelder, & Meunier, 1998; Hallé & Best, 2007; Moreton, 2002).

The present study focuses on the lexical context effect first demonstrated by Ganong (1980). Ganong presented listeners with three voice onset time (VOT) continua (/b/-/p/, /d/-/t/, and /g/-/k/). The following context formed a word at one endpoint of the continuum but not the other—for example, with the /d/-/t/ continuum: *–ash*, where /d/ makes a word (*dash*) but /t/ does not (**tash*), or *–ask*, where /t/ makes a word (*task*) but /d/ does not (**dask*). Listeners tended to categorize ambiguous sounds along the continua so as to make a word in the context provided. This finding has since been replicated many times and is now widely known as the "Ganong effect" (see Pitt & Samuel, 1993, for additional results, a review, and a meta-analysis of results to that date).

The Ganong effect has been interpreted differently in 'interactive' and 'autonomous' models of speech perception. In the interactive TRACE model (McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006), the Ganong effect is attributed to feedback from lexical activation to perceptual processing of the phonemes that a word contains. If one of the candidate phonemes receives facilitatory lexical feedback, and the other does not, the activation of the former will increase in strength, and its increased activation will reduce the activation of the latter through inhibitory connections between phonemes. For example, if a sound that is ambiguous between /d/ and /t/ is followed by *–ash*, activation of the word *dash* will increase activation of /d/, which will in turn inhibit activation of /t/. The back-and-forth flow of activation between the phonemic and lexical levels then continues (e.g., the increased activation of /d/ leads to still stronger activation of *dash*), further increasing activation of the candidate phoneme that has lexical support and further inhibiting the one that lacks it.

In autonomous models (e.g., Cutler & Norris, 1979; Norris, McQueen, & Cutler, 2000), on the other hand, phonemic activation itself is uninfluenced by linguistic knowledge. In the Merge model (Norris et al., 2000), the Ganong effect is attributed to a combination (or merging) of the effects of phonemic activation and lexical activation in the service of making the phonemic decisions required in categorization experiments. Lexical candidates are activated by phonemic activation, but this activation does not feed back to influence phonemic activation. Instead, lexical and phonemic activation both feed forward to nodes responsible for explicit phonemic decisions.

Because lexical influence on phonemic decisions depends on a feedback loop between the phonemic and lexical levels in TRACE, and because it takes time for activation to flow between levels, TRACE predicts that lexical involvement in phonemic decisions should take time to emerge, and should gradually and monotonically increase in strength. McClelland and Elman (1986; see Fig-

This article was published Online First August 15, 2016.

John Kingston, Department of Lingustics, University of Massachusetts, Amherst; Joshua Levy, Department of Psychological and Brain Sciences, University of Massachusetts, Amherst; Amanda Rysling, Department of Lingustics and Department of Psychological and Brain Sciences, University of Massachusetts, Amherst; Adrian Staub, Department of Psychological and Brain Sciences, University of Massachusetts, Amherst.

The order of authors is alphabetical.

Correspondence concerning this article should be addressed to John Kingston, Department of Lingustics, University of Massachusetts, N412 Integrative Learning Center, 650 North Pleasant Street, Amherst, MA 01003. E-mail: jkingston@linguist.umass.edu

ure 6) simulate the Ganong effect with an ambiguous sound on the /p/-/b/ continuum, where the following context is either *–lug* (where /p/ makes a word but /b/ does not) or *–lush* (where /b/ makes a word but /p/ does not; the word *plush* is not included in the TRACE lexicon). The simulation shows that the final segment does not influence activation of /p/ or /b/ immediately, but only after a period equivalent to the duration of almost two more segments has elapsed. This influence then grows gradually. The lexical influence is still increasing in strength at a point equivalent to seven segments' duration after the onset of the final segment. (McClelland and Elman's display of simulation results only continues to that point, but it is clear that in fact the lexical influence would continue to grow even after this point.) In sum, the mechanism that TRACE uses to account for the Ganong effect requires that it be a slow-starting, gradually developing, and monotonically increasing effect. While the effect's observed time course will depend on specific model parameters, the basic time course prediction is intrinsic to any model on which the effect arises through temporally extended feedback loops, because each increase in a word's activation increases its constituent phonemes' activations, which in turn increase that word's activation.

Initial results from experiments in which response time (reaction time (RT)) was recorded together with the subject's categorization seemed to support the prediction that the Ganong effect develops gradually. Fox (1984) found that for initial consonant targets, the size of the lexical effect on response proportions was larger for slow RTs than for fast RTs. McClelland and Elman (1986) argue that TRACE predicts Fox's finding if fast RTs reflect trials on which the subject made a phonemic decision before lexical feedback could influence at the phoneme level, but slow RTs reflect the full operation of the feedback loop between phonemic and lexical activation. Miller and Dexter (1988) obtained a pattern similar to Fox (1984), and Connine and Clifton (1987) found that categorizing ambiguous stimuli was faster when the response category formed a word with its context than when it formed a nonword, but that word and nonword categorization did not differ in RT for less ambiguous stimuli near continuum endpoints. This pattern was also argued to support TRACE, as lexical feedback should speed responses to ambiguous intermediate stimuli, for which responses are otherwise relatively slow (Pisoni & Tash, 1974), but should not affect responses to endpoint stimuli, whose unambiguous acoustic information alone can lead to a fast response.

McQueen (1991) argued that interactive and autonomous models make clearly distinct RT predictions only when the target sound is word-final. In this case, the TRACE model predicts that the Ganong effect should be especially strong for slow responses. McQueen pointed out that for word-final phonemes, however, the autonomous model of Cutler, Mehler, Norris, and Segui (1987) predicts a stronger lexical preference in fast responses because all information that can influence lexical activation has already been heard. McQueen found the predicted stronger lexical preference for fast responses for word-final targets. McQueen, Cutler, and Norris (2003) replicated this pattern, but this time found stronger lexical preferences in fast responses both when the target sound was word-initial and when it was word-final.

The RT data are not entirely consistent, however. Pitt and Samuel (1993, Experiment 1) found that the Ganong effect grew with RT for two initial voicing continua but instead shrank with RT for a third initial manner continuum, and found that word responses were faster than nonword responses to ambiguous but not endpoint stimuli only for the same two initial voicing continua where the Ganong effect grew with RT. For the third manner continuum, word responses were faster than nonword responses for endpoint stimuli, but not ambiguous stimuli (cf. Connine & Clifton, 1987). For final consonant targets, Pitt and Samuel (1993, Experiment 2) replicated McQueen's (1991) finding of a stronger Ganong effect in fast responses, but only for voicing and not the manner continuum, and they showed faster word than nonword responses to unambiguous but not ambiguous stimuli for both continua.

This literature is therefore inconclusive. Equally important, patterns of behavioral RTs support only indirect inference as to precisely when, in the course of a trial, lexical activation influences phonemic activation. In these experiments, most behavioral responses occur hundreds of milliseconds after the offset of the auditory stimulus itself. For example, in McQueen et al.'s (2003) Experiment 1, mean RT was 958 ms, while the stimuli themselves were approximately 500–600 ms in duration; and in the initial consonant condition, the critical phoneme occurred much earlier than this. In their Experiment 2, they put participants under severe pressure to respond quickly and obtained a shorter mean RT of 737 ms, a value still longer than the stimuli. Despite these shorter RTs, the Ganong effect still remained strongest in the fast responses and shrank dramatically or disappeared entirely in the slow ones. These experiments assume that the time at which a behavioral response is made (i.e., the RT) is at least a monotonic function of the time at which a phonemic decision was made. This is a reasonable, but unnecessary, assumption.

Our experiments use a method that can assess more directly when, over the course of a trial, lexical knowledge influences phonemic activation. These experiments monitored subjects' eye movements to visually presented response options during a categorization task. Because eye movements executed during language comprehension tasks are rapid and relatively automatic, and because the earliest eye movements precede overt behavioral responses, such as button-presses, by many hundreds of milliseconds, they may reveal fine details about the timing of lexical effects on speech sound categorization.

In particular, we can address two questions. First, how rapidly does lexical knowledge influence eye movements relative to when the stimulus' acoustic properties could, in principle, provide evidence for a particular word? In the example stimuli from Ganong (1980), when the target sound is ambiguous between /d/ and /t/, and the context is *–ash* or *–ask*, lexical knowledge cannot be applied to the categorization decision any earlier than when the third segment becomes audible. At this moment, eye movements to potential response options might begin to diverge immediately in the two contexts. On the other hand, there may be a delay between when this segment's acoustic properties are heard and when it influences eye movements. The TRACE model appears to predict such a delay (McClelland & Elman, 1986, Figure 6), because it takes time for word activation to feed back on phoneme activation and start the mutual feedback loops. The question of how rapid an influence can be (i.e., whether an eye movement effect would count as immediate or delayed) could be answered using a previously established estimate of how long it should take an effect of a linguistic manipulation to appear in the eye movement record. A 200 ms figure is often attributed to Matin, Shao, and Boff (1993), but using an arguably more appropriate method, Altmann (2011) arrived at an estimate of no more than 100 ms. In the present experiments, however, we can address this question simply by examining how rapidly the acoustics of the target sound (/d/-/t/ in the current example) influence eye movements, and by asking whether the lexically informative segment influences eye movements as rapidly as the target sound itself, relative to their respective onsets.

The second question is whether the influence of lexical knowledge continues to grow throughout the trial. TRACE predicts that once lexical feedback begins to influence phonemic activation, this influence should continuously increase in size through feedback loops between the lexical and phonemic levels. Some behavioral results (e.g., McQueen et al., 2003), however, suggest that the Ganong effect is either very small or absent altogether when RT is very slow. McQueen et al. (2003) have argued that Merge can accommodate this finding, as there is no feedback loop causing a gradually increasing lexical influence on phonemic activation. However, Merge apparently does not directly predict a diminishing lexical effect over time, as found by McQueen et al. (2003). Our experiments permit us to assess whether the influence of a lexically informative phoneme gradually increases once this phoneme is heard, or whether this is primarily or entirely an early influence that diminishes with time. We introduce a data analytic method that is designed to determine when an effect on eye movements begins to wane.

Previous research has demonstrated that eye movements reveal details of the process of categorizing a speech sound, beyond what can be learned using traditional behavioral paradigms. McMurray, Tanenhaus, and Aslin (2002) and McMurray, Aslin, Tanenhaus, Spivey, and Subik (2008) presented listeners with a /b/-/p/ VOT continuum as the initial consonant in a minimal pair (e.g., *bear* vs. *pear*). The task was to click with a computer mouse on the corresponding image. When the VOT was close to the category boundary, listeners looked significantly more often at the image depicting a word whose initial consonant was on the other side of the category boundary than when the VOT was farther from the category boundary. This was observed even when analysis was restricted to those trials during which the listener clicked on the image corresponding to the near side category. For example, listeners were relatively likely to look at the picture of the bear before clicking on the pear when the VOT was relatively short for a /p/, compared to when the VOT was more typical of a /p/. McMurray et al. interpreted these looks to the competing visual stimulus as evidence that the /b/-/p/ distinction is in fact a gradient rather than categorical one.

Other studies anticipate the present work in using eye movements to assess the detailed timing of effects on speech sound categorization. McMurray, Clayards, Tanenhaus, and Aslin (2008) presented VOT (/b/-/p/) and transition duration (/b/-/w/) continua to listeners before short and long vowels. Listeners judge a given VOT and transition duration as shorter, and thus more indicative of /b/, when the following vowel is long (Miller & Liberman, 1979; Miller & Volaitis, 1989). McMurray et al. (2008) found that both VOT and transition duration influenced listeners' early looks to images, but vowel duration only influenced later looks. This asynchrony corresponds to when each kind of information became available in the signal.

Mitterer and Reinisch (2013) used eye movements to investigate how early in the process of phonemic categorization effects of perceptual learning can emerge. In a between-subjects design, subjects first were presented with a fricative that was ambiguous between /f/ and /s/ in lexical contexts that caused it to be heard either as an instance of /f/ or as an instance of /s/. The manipulation of initial exposure influenced eye movements as rapidly as did the acoustics of the target sound itself in a subsequent categorization task.

Our experiments use a variant of the visual world paradigm similar to that of Mitterer and Reinisch (2013), in which text is displayed rather than pictures (McQueen & Viebahn, 2007). Because a goal of the present experiments was to mimic the task used in the classic Ganong-effect studies (e.g., Ganong, 1980; Pitt & Samuel, 1993) as closely as possible, however, we used letters rather than words as the response options. Listeners' tasks were to click on the letters corresponding to the sounds they heard in the stimulus.

In our first experiment, we used an initial target in a one-syllable stimulus (a sound from an /s/-/f/ continuum), as in the original Ganong (1980) study. In addition to contexts in which /s/ makes a word but /f/ does not (*-ide*) or in which /f/ makes a word but /s/ does not (*-ile*), we included a context in which neither category makes a word (*-ime*). As in the Ganong stimuli, it is not until the third and final segment that it is possible, in principle, for the listener to apply lexical knowledge. We compare the timing with which the third segment's identity influences eye movements against the timing with which the initial segment's acoustics influences eye movements. In our second experiment, the target sound was the nuclear vowel in a one-syllable stimulus, and the immediately following consonant conveyed which, if any, vowel category made a word. In this experiment, listeners could, in principle, apply their lexical knowledge almost immediately upon presentation of the target sound.

The present work also demonstrates a simple, but novel, method for the analysis of eye movement data from visual-world type tasks. This analysis is especially suitable for relatively precise temporal localization of the onset and offset of an effect of an experimental manipulation. It focuses on the location of eye fixations that begin in each successive time window during a trial, effectively distinguishing between effects on current eye movements and effects that are due to continuing fixations begun in previous time windows. This novel analysis is presented alongside the traditional analysis.

We thus expect that the timing of eye movements reflects the temporal evolution of the influence of lexical knowledge on categorization. It is also necessary, however, to ask when the stimulus provides acoustic evidence that could, in principle, activate a word. Accordingly, we ran gating experiments (Smits, Warner, McQueen, & Cutler, 2003) using the same stimuli with new groups of listeners to determine how much of each stimulus must be heard to produce a Ganong effect. The eye tracking experiments are presented as Experiments 1a and 2a, and the corresponding gating experiments as Experiments 1b and 2b.

Experiment 1a

Method

Subjects. The 22 subjects were undergraduate students at the University of Massachusetts Amherst, who earned psychology course credit for their participation and were naïve to the purpose of the experiment. All were monolingual native speakers of English with normal or corrected-to-normal vision. None reported any history of speech or language disorder. They all gave informed consent before participating.

Materials. The target sounds in this experiment consisted of nine steps from a 20-step /s/-/f/ continuum, from an unambiguous /s/ (Step 1) to an unambiguous /f/ (Step 20). The nine steps were 1, 5, 7, 9, 11, 13, 15, 17, and 20. These steps were used in order to have both unambiguous /s/ and /f/ anchors and to sample the ambiguous portion of the continuum densely. The target was heard as the first segment in a monosyllable. Three different syllable rimes were used: *-ide* (/aɪd/), *-ile* (/aɪl/), and *-ime* (/aɪm/). With the first rime, /s/ makes a word but /f/ does not (*side* vs. **fide*). With the second rime, /f/ makes a word but /s/ does not (**sile* vs. *file*). With the third rime, neither onset makes a word (**sime* vs. **fime*).[1]

These materials were chosen so as to minimize the influence of lexical statistics on the likelihood of listeners responding "f" versus "s". Table 1 lists the word-form frequencies of *file* and *side*, their frequency-weighted neighborhood densities, and the transitional probabilities between the first segment /f/ or /s/ and the following nucleus /aɪ/, computed from the Subtlex corpus (Brysbaert & New, 2009; as obtained from the Irvine Phonotactic Online Dictionary, Vaden, Halpin, & Hickok, 2009).

The stimuli were made by editing waveforms of syllables produced by a male speaker of American English. The rimes *-ide*, *-ile*, and *-ime*, were recorded with the onset /h/—that is, *hide*, *hile*, and *hime*. This sound was chosen because there are no formant transitions between /h/ and the following vowel, and thus no coarticulatory information about the place of articulation of the preceding consonant at the beginning of the nucleus. The /h/ was spliced off and members of the /s/-/f/ continuum spliced in. The /s/ and /f/ were recorded before /aɪ/, so their acoustics would not be affected by long-distance coarticulation with any final consonant. The tokens used were those from a larger set that exhibited the largest differences in energy between a lower frequency 100–4000 Hz band and a higher frequency 4000–8000 Hz band characteristic of these two fricatives: In the /s/, energy in the higher band was 15.99 dB greater than that in the lower band, while in the /f/, energy in the lower band was 5.92 dB greater than that in the higher band. The continuum was made by mixing tokens of that /s/ and /f/ in linear incrementally complementary proportions, from 1.0 /s/ + 0.0 /f/ to 0.0 /s/ + 1.0 /f/.

Table 2 lists the duration of each interval in ms. The boundary between the nucleus and final consonant was determined by careful listening and visual inspection of spectrograms. The spectrograms in Figure 1 show that the formant trajectories in the vowels begin to visibly differ between the three rimes roughly 350 ms after stimulus onset. See also the results of the gating experiment, Experiment 1b, for an empirical test of when these final consonants became audible.

Table 1
Word Frequency (Counts Per Million), Frequency-Weighted Neighborhood Density, and Forward [p(Nucleus|Consonant)] and Backward [p(Consonant|Nucleus)] Transitional Probabilities for File and Side

	Frequency	Neighborhood Density	Forward Transitional Probability	Backward Transitional Probability
File	44.04	2975.63	.0584	.0616
Side	200.92	2065.53	.0241	.0445

Table 2
Durations (in ms) of Stimulus Intervals in Experiments 1a, b

Stimulus	Fricative	Nucleus	Final Consonant
s/fide	150	225	195
s/file	150	215	220
s/fime	150	225	195

Each subject heard 190 stimuli using each of the three rimes, for a total of 570 experimental trials. The 190 stimuli with each rime consisted of 10 occurrences of Steps 1 and 20; 20 occurrences of Steps 5, 7, 15, and 17; and 30 occurrences of Steps 9, 11, and 13. The 570 trials were presented in a random order to each subject following presentation of 12 practice trials, in which each of the six endpoint stimuli was presented twice with correct answer feedback.

Procedure. The subject's task was to respond whether /f/ or /s/ was heard on each trial by clicking with a mouse on a visually presented uppercase letter F or S. The letters were presented in 36-point Monaco font, centered 192 pixels above (F) and below (S) the center of a 19-inch monitor with a resolution of 1024 × 768 pixels, placed 70 cm from the participant. Auditory stimuli were played over Sennheiser HD 280 64Ω headphones. Stimulus presentation was controlled using the Experiment Builder (SR Research, Toronto, Ontario, Canada) software package.

Each trial began with a red circle appearing in the center of the screen. The subject was required to visually fixate this circle to initiate the trial, which allowed the experimenter to perform an eye fixation drift correction and to check calibration. When the subject fixated the circle, it changed color to green and the auditory stimulus was played. The subject clicked on the letter corresponding to the sound that he or she heard. The subject was then required to click on an X that appeared at the location of the center circle to terminate the trial, which reset the location of the computer mouse.

Eye movements were recorded using a remote desktop camera connected to an Eyelink 1000 tracker (SR Research, Toronto, Ontario, Canada). Gaze location was sampled at a rate of 500 Hz. Viewing was binocular, but only the right eye was tracked. Before the experiment began, tracking was calibrated by having the subject fixate a sequence of 13 dots on the computer screen. A calibration was accepted if average error was less than 1 degree of visual angle. Calibration was repeated as necessary during the experiment, as indicated by the between-trial drift-correction procedure. The experiment took approximately one hour.

[1] The third segment of the *-ile* context does not rule out longer words beginning with /s/ (e.g., *silo*, *silent*), nor does the third segment of the *-ime* context rule out longer words beginning with /s/ (e.g., *simultaneous*). We hypothesized that within the context of an experiment deploying an extremely limited set of stimuli, all of which were one syllable, these potential lexical hypotheses may not actually be entertained by subjects, so that the third segment would effectively rule out a word response in these cases. The data apparently support this conjecture. Previous evidence shows that listeners are sensitive to differences in phonetic detail that distinguish one-syllable words from initial syllables of longer words that consist of the same segments (e.g., Salverda, Dahan, & McQueen, 2003).

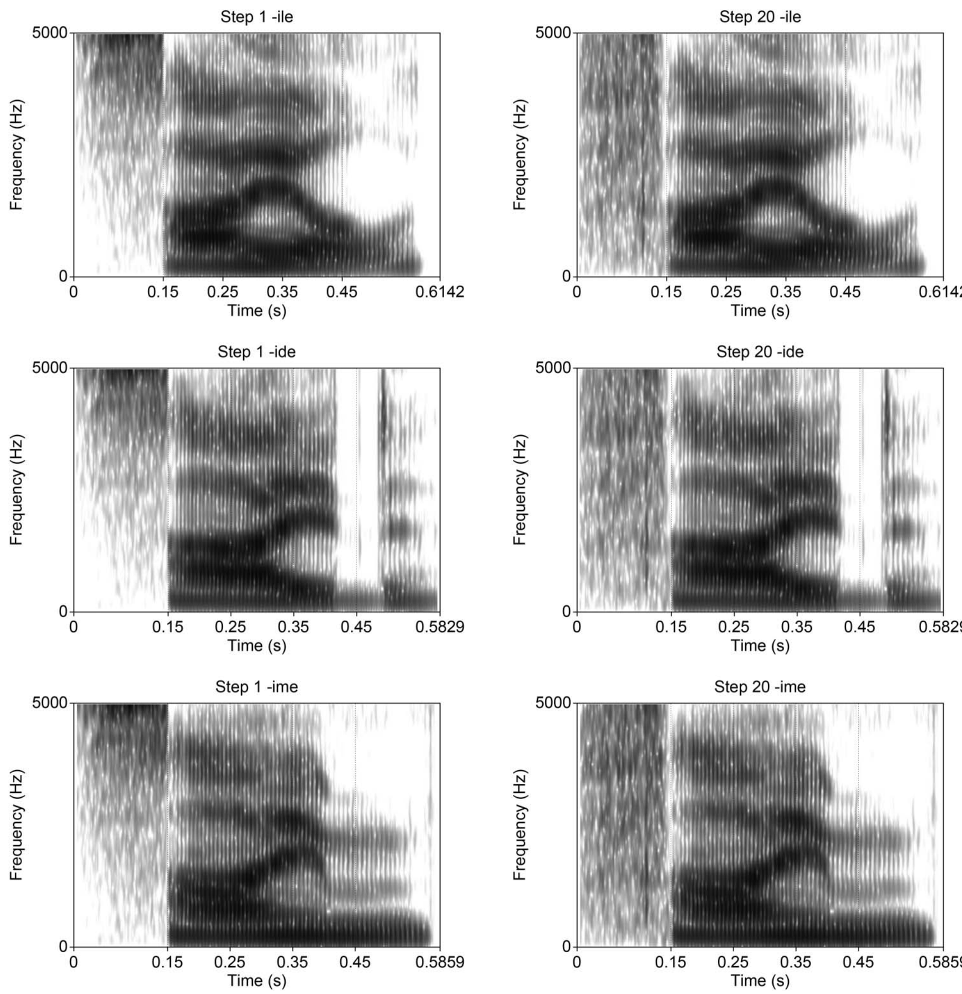

Figure 1. Spectrograms of Steps 1 (left) and 20 (right) of the stimuli with *-ile* (top), *-ide* (middle), and *-ime* (bottom) rimes. The labeled tick marks at the bottom of each spectrogram show the locations of the gates used in Experiment 1b.

Results

The behavioral results are shown in Figure 2. As expected, the proportion of 'F' responses increased monotonically from the /s/ to the /f/ endpoint of the continuum, following the logistic function that is typical of phonetic categorization experiments. Listeners clearly responded 'F' much more often in the *-ile* context than the *-ime* context, especially at the continuum midpoints. The preference for 'F' responses extended to the /s/ end of the continuum, where the *-ile* context elicited some 'F' responses to an unambiguous /s/ sound. The difference in the proportion of 'F' responses between the *-ide* and *-ime* contexts is far less dramatic, but is in the predicted direction, with fewer 'F' responses before *-ide* than before *-ime*.

We computed a mixed effects logistic regression model of the probability of an 'F' response using the glmer() function from the lme4.0 package (Bates, Maechler, Bolker, & Walker, 2014) for the R statistical programming environment (R Core Team, 2014). Fixed effects included stimulus step, centered on 10.5 (the midpoint between the endpoints of 1 and 20); context, using treatment contrasts with the *-ime* context coded as the reference level; and the interaction of these variables. Random effects of subject were included on intercepts and slopes for each of the main effects, but random effects of subject on interaction slopes had to be omitted for the models to converge. Parameter estimates are shown in Table 3. The negative intercept reveals a slight overall bias against responding 'F.' 'F' responses increased significantly with step. The probability of an 'F' response was greater in the *-ile* context than in the *-ime* context, and was also significantly lower in the *-ide* context than in the *-ime* context, though the latter effect was much smaller. Neither interaction reached significance, though one was marginal.

An eye fixation was coded as falling on F or S if the fixation was located more than 66 pixels above or below the center of the screen, respectively. This criterion was based on the spatial distribution of eye fixations, which were concentrated on the locations of the letters or on the central circle.

Figure 3a displays fixation proportions on F over the course of a trial as a function of continuum step. Continuum step begins to influence the probability of a subject fixating F before 400 ms after stimulus onset. Figure 3b displays the proportion of F fixations over time as a function of context. (Endpoint trials, where context had little effect, are not included in this figure.) The divergence here appears around 500 ms, when the probability of fixating F begins to increase noticeably in the *-ile* context compared to the other two contexts.

We first used a standard method for determining when the experimental manipulations affected the eye movement record. For each 100 ms time bin after the onset of the stimulus from 100 ms to 1800 ms, a logistic regression model was fit to the probability of the eyes fixating on F during that bin as a function of continuum step, context, and their interactions. We excluded endpoint trials, because as expected, there were small or nonexistent behavioral effects of the context manipulation with these stimuli. For each bin, the model's fixed and random effects were specified in the same way as for the model of the behavioral results.

Figure 4a shows the critical results graphically for each time bin, in the form of z scores associated with the three main effects (continuum step, *-ide* vs. *-ime*, and *-ile* vs. *–ime*). In the first bin shown, 100–200, neither continuum step nor lexical context significantly influence fixations. These absences rule out an early bias to look at F. However, by the 200–300 ms bin, continuum step has a significant effect (based on a criterion of $|z| > 2$) on the probability of fixating F.[2] This effect gets much larger over the next several hundred milliseconds, and it continues to the end of the analyzed time period. Compared to the effect of continuum step, the effect of context is somewhat delayed. Fixations on F are not significantly more likely in the *-ile* context than the *-ime* context until the 300–400 ms bin ($z = 2.009$), but this difference is then highly significant throughout the remainder of the trial. Even though fixations on F are less likely in the *-ide* than the *-ime* context, this effect reaches significance in only a few bins, beginning in the 500–600 ms bin, and in these cases just barely. This is not unexpected, given that the corresponding behavioral effect was very small, and given the eye movement patterns shown in Figure 3. Figure 4b shows the z scores corresponding to the interaction effects. While these show some variability over time, no trend continues over multiple time bins, suggesting that the occasional barely significant result may reflect the expected Type I error rate. The absence of clear interaction effects on eye movements is also expected, given the lack of interaction effects in the behavioral data.

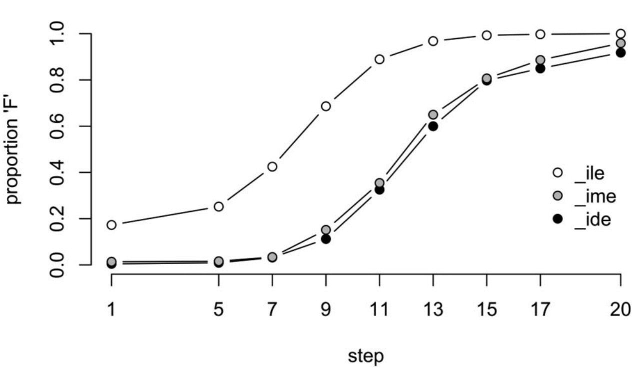

Figure 2. Experiment 1a response proportions as a function of continuum step and context.

Table 3
Parameter Estimates, Standard Errors, z Values, and p Values From a Mixed-Effects Logistic Regression Model of the Probability of an 'F' Response in Experiment 1a

Parameters	Estimate (β)	Std. Error	z Value	p Value
Intercept	−1.039	.248	42.00	<.001
Step	.718	.062	11.65	<.001
/aɪd/ vs. /aɪm/	−.464	.202	−2.29	.022
/aɪl/ vs. /aɪm/	3.269	.311	10.52	<.001
Step × /aɪd/ vs. /aɪm/	.052	.028	1.84	.067
Step × /aɪl/ vs. /aɪm/	−.020	.028	−.71	.478

An additional analysis then focused on when each manipulation affected the movement of the eyes, rather than the location of the eyes. In each time window, this analysis was restricted to trials on which the subject *initiated* an eye fixation on either F or S in that time window.[3] The dependent measure in mixed-effects logistic regression models for each 100 ms time window was the probability that this new fixation fell on F as opposed to S. The analysis also excluded trials on which a new fixation was preceded by another fixation on the same letter; this has the effect of counting only the first fixation of a run of two or more consecutive fixations to the same letter.

The traditional analyses and this new analysis may diverge with respect to the timing of effects, especially with respect to when an effect ends. Eye fixations average between 200 and 300 ms in duration, though the distribution is right-skewed and includes fixations as long as 1 s in duration (see Staub, Abbott, & Bogartz, 2012, for distributions of fixation durations in a visual world task). As a result, the locations of fixations that began in a time bin previous to the one under analysis will determine, in part, the probability that the eyes are now fixating a given stimulus. Therefore, a given factor can have a statistically significant influence in time window n even if that factor no longer has any influence on the movement of the eyes in window n, but had a pronounced effect in windows n-1, n-2, and so forth. Persistence from previous time windows cannot result in a significant effect in this novel analysis, however. By excluding fixations to a letter that began in previous time windows, comparisons within a single time window provide a better estimate of when particular stimulus properties influence the proportion of F fixations.

[2] Correcting for up to 17 multiple comparisons, the z score corresponding to an α of 0.05 is 2.76.

[3] We excluded new fixations that fell on the central region of the display rather than a letter, for two reasons. The first reason is theoretical: the new analysis focuses as specifically as possible on how the experimental manipulations influenced the decision to fixate one letter versus the other. The second reason is practical: in an analysis that examines the probability that any new fixation in a given time window falls on F (that is, including in the denominator fixations that are on the central region, rather than a letter), it is difficult, if not impossible, to make meaningful comparisons across time windows. This is because the proportion of new fixations that fall on the central region, rather than on one of the letters, is much greater earlier in the trial than later. Put differently, the overall proportion of new fixations that fall on a letter, as opposed to the central region, increases over the course of the trial in any visual world experiment, and what is of interest are the effects of the experimental manipulations on the location of these fixations.

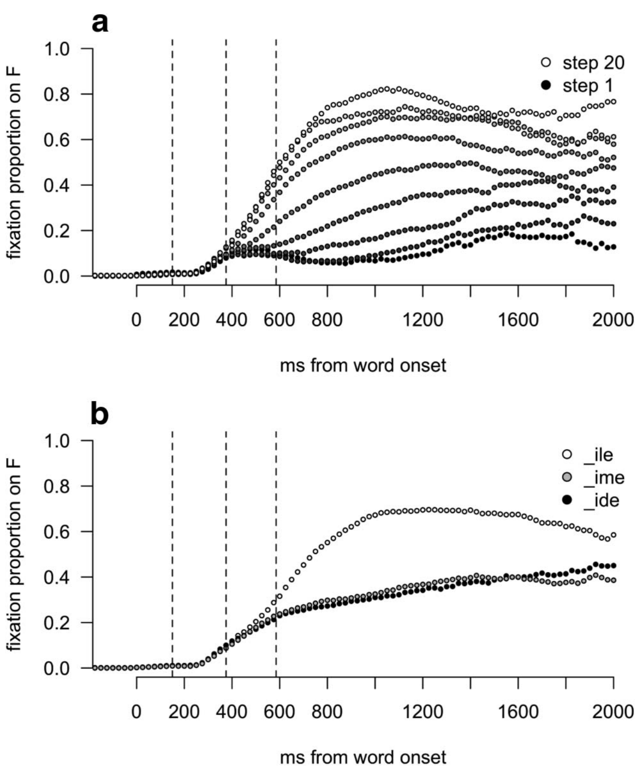

Figure 3. Experiment 1a average fixation proportion on F over time as a function of (a) continuum step and (b) context. Progressively lighter shades of gray between black, Step 1, and white, Step 20, in (a) represent successive steps along the /s/-/f/ continuum. Endpoint steps are excluded from the latter figure. Vertical lines at 150 ms, 375 ms, and 585 ms represent the offset of the fricative, nucleus, and final consonant, respectively, based on the latest of the offsets in Table 2.

Model specification was the same as for the traditional analysis. The results are shown in Figure 5. (The earliest bin shown is 200–300 ms because there were too few new fixations in the 100–200 ms bin for the model to converge.) The estimate of the timing of the effect of continuum step is consistent with the earlier analysis, except that the effect's size now clearly declines over the course of the trial, so that after about 1500 ms the effect is no longer consistently significant. This discrepancy suggests that in the later time windows, effects of continuum step on the probability of fixating on F are largely due to persistent earlier fixations, rather than to new fixations. The effect of context declines similarly over time in this analysis. Finally, the two analyses differ slightly in when fixations on F first increase in the *-ile* context compared to the *-ime* context: The onset is slightly later in the new analysis than the old. The interactions do not noticeably differ from the preceding analysis.

Figures 6a and 6b show the proportion of new fixations, of those that are on a letter, that fall on F in the 400–500 ms and 500–600 ms bins, respectively, as a function of step and context. Consistent with the statistical model, in the 400–500 ms bin there is a very clear effect of step, but not context; there are numerically more fixations on F in the *-ile* context than in the other contexts only for three of the seven steps. In the 500–600 ms bin, however, there is also a very clear effect of context, with more F fixations in the *-ile* context than in the other two contexts at every continuum step. The proportions of fixations on F in the *-ide* context do not differ systematically from those in the *-ime* context in either bin.

The influence of lexically supportive contexts on fixations emerged statistically at the same time when *-ile* and *-ide* were directly compared, first in the 400–500 ms bin in the traditional analysis and in the 500–600 ms bin in the novel analysis.

Discussion

As expected, subjects tended to categorize an ambiguous sound from the /s/-/f/ continuum so as to form a word. This tendency was very pronounced in the comparison of the /f/-biased *-ile* context with the neutral *-ime* context, but was much less pronounced, though still significant, in the comparison of the /s/-biased *-ide* context with the neutral context. This asymmetry has also appeared in previous behavioral experiments in our laboratories using the same contexts (Rysling, Kingston, Staub, Cohen, & Starns, 2015), as well the gating experiment, Experiment 1b. The reasons for the asymmetry are not clear, but this issue is unrelated to this article's main question.

The main question is when in the course of a trial lexical knowledge exerted its influence, relative to when it could first exert an influence—that is, the onset of the third segment of the stimulus. The eye movement analysis showed, first, that the acoustics of the target sound influenced eye movements within the first 400 ms of the onset of the stimulus. The context did not influence eye movements until somewhat later; precisely how much later

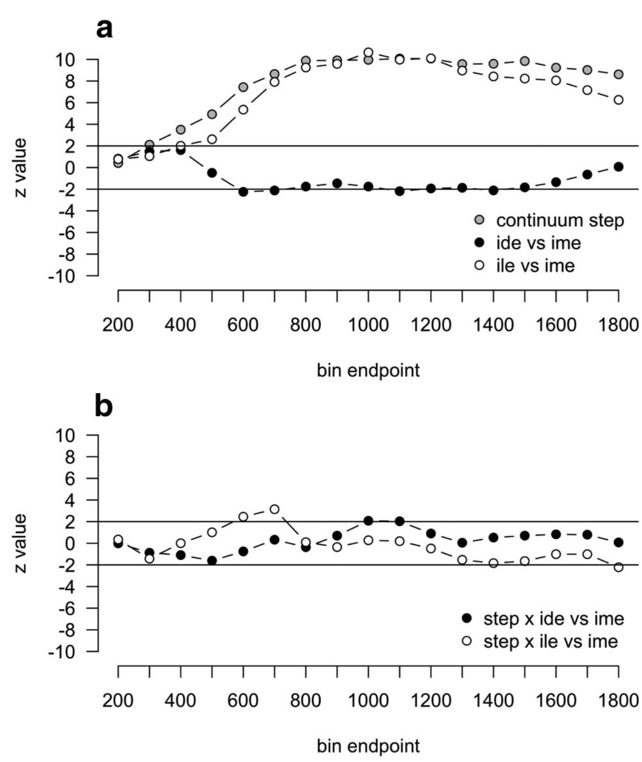

Figure 4. Experiment 1a *z* scores from a mixed effects model of eye movement data for (a) the three main effects and (b) the two-way interactions.

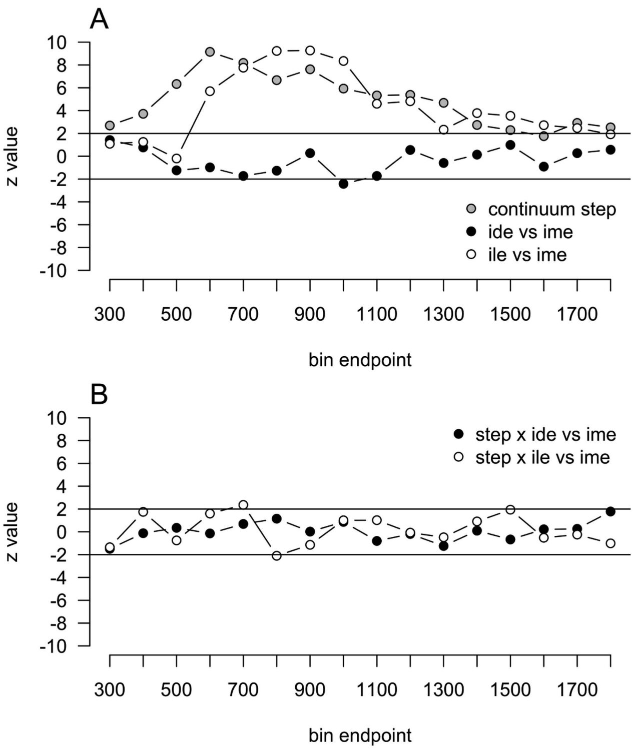

Figure 5. Experiment 1a *z* scores from a mixed effects model of eye movement data for (a) the three main effects and (b) the interaction effects, based on the alternate analysis method described in the text.

depends on the details of the statistical analysis, but certainly by the 500–600 ms time bin. Relative to the onset of the respective segments, however, this pattern suggests that the third segment's effect was at least as rapid as the first segment's effect. The third segment followed the first by at least 200 ms, and there is no more than a 200 ms difference (perhaps less) in the point at which each of these manipulations influenced eye movements.

It is simple to explain the discrepancy between the two analysis methods regarding the onset of the effect of context, which is significant as early as the 300–400 ms bin in the traditional analysis but not until the 500–600 ms bin in the analysis of new fixations. The proportion of new fixations on F exhibits a very small numerical advantage in the *-ile* context starting shortly after the onset of the trial. Thus, the overall difference between contexts gradually increases, reaching significance in the 300–400 ms bin. However, in each individual bin the effect of context on the location of *new* fixations is very small indeed, until the 500–600 ms bin, when this effect is quite large.

The two analysis methods answer different questions, just as they should: First, how do fixations to a particular target accumulate or persist, and second, what prompts a new fixation to a target? It may generally be useful to conduct both analyses and to compare the verdicts they deliver. A drawback of the traditional analysis is that it has the potential to somewhat arbitrarily identify a specific time bin in which an effect becomes significant, when in fact the fixation proportion difference between conditions has been gradually increasing over time. The analysis of new fixations alone can miss gradually developing effects on fixation locations, as it is a "memoryless" analysis in which data in each time window are unaffected by eye movements in previous windows. Neither analysis is therefore without flaws. We return to these differences in the General Discussion.

In the traditional analysis, the influence of both continuum step and context extended well beyond 1 s after stimulus onset. These effects were less persistent in the analysis of new fixations, as would be expected given that the locations of fixations in the one-time bin do not carry over to influence the statistical analysis of subsequent bins. Still, the persistence that is demonstrated in both analyses is surprising, given that the locations of fixations beginning hundreds of milliseconds after the offset of the stimulus are being influenced by these manipulations. Most important for present purposes, however, is that the effect of lexical knowledge activated by the third segment's identity reaches its maximum relatively quickly, and then gradually declines over the rest of the trial. Thus, the data do not support the idea that the influence of lexical knowledge gradually increases. If anything, this effect declines more rapidly than does the effect of target sound acoustics.

We turn next to the companion gating experiment, which provides a measure of when listeners could have heard acoustic evidence of the final consonants in the stimuli and thus when words could first have been activated.

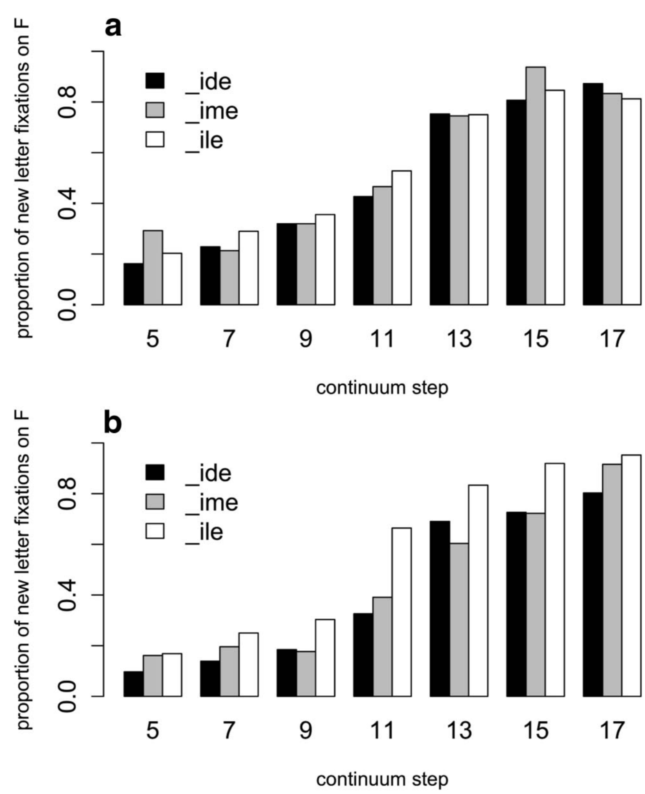

Figure 6. Experiment 1a proportion of new fixations on a letter in the 400–500 ms bin (a) and in the 500–600 ms bin (b) that fall on *F*, as a function of context and step.

Experiment 1b

The goal of Experiment 1b was to determine empirically when listeners could first have heard acoustic properties of the final consonants /l/, /d/, or /m/, which could in turn have activated the words *file*, *side*, or no word, respectively. By evaluating lexical bias across initial fragments of the stimuli that vary incrementally in how long they last, we can estimate when these acoustic properties become audible. Inspection of the spectrograms in Figure 1 of the rimes *-ile*, *-ide*, and *-ime* and careful listening suggests that this point was about 350 ms after stimulus onset for all three contexts. To assess this estimate's accuracy, we tested a new group of listeners in a gating experiment using the procedures for gating stimuli described in Smits et al. (2003).

Methods

Subjects. Participants were 20 undergraduates at the University of Massachusetts, Amherst. All were adult native speakers of English who reported no exposure to any language other than English before the age of 6 and no hearing or speaking disorders. They gave informed consent and were compensated with course credit.

Materials. The endpoints, Steps 1 and 20, and the most frequently presented intermediate Steps, 9, 11, and 13, from the /s/-/f/ continuum in Experiment 1a were gated at 150, 250, 350, 450, and 705 ms. The longest gated fragment (henceforth a "gate") lasted longer than the stimuli. The locations of the tick marks at the bottoms of the spectrograms in Figure 1 show that the shortest gate (150 ms) corresponded to the end of the initial fricative, before the following vowel began, and the next gate (250 ms) occurred well into the vowel, but before the formants begin to change in ways that would convey the identity of the following consonant. By the gate at 350 ms, the rimes differ acoustically from one another, and by the gate at 450 ms, the listeners would have heard substantial acoustic information about the identity of the final consonant. Finally, at the longest gate, 705 ms (not shown in the figure), they would have heard the entire stimulus.

So that cutting the stimulus off abruptly did not introduce spurious percepts of a following voiceless stop, each gate was followed by a square wave with an F0 of 500 Hz that lasted until 705 ms. The square wave's RMS amplitude was −22 dB less than the most intense 50 ms interval of the preceding gate. The gate was ramped down to 0 and the beginning of the following square was ramped up to its maximum amplitude by multiplying their final and initial 5 ms, respectively, by a raised cosine window. The gate and square wave were then spliced together with 5 ms of overlap.

Procedure. Listeners were first presented with 20 training trials in which the endpoints, Steps 1 and 20, were presented twice in the no lexical bias *-ime* context at each of the five gates with correct answer feedback following each trial. Training was followed by 3 blocks of test trials in which Steps 1, 9, 11, 13, and 20 were presented in a ratio of 1:4:4:4:1 in each of the three contexts at each of the five gates. In all, there were 210 trials per test block and 630 test trials per listener. Correct answer feedback was still given to the endpoint stimuli during the test blocks, but not to the intermediate stimuli. In both training and test blocks, stimuli were presented in a different random order to each participant. Participants took brief self-timed breaks one third and two thirds of the way through each test block (every 70 trials), and longer breaks between blocks.

At the same time as the stimulus began to play on each trial, the response prompts 'S' and 'F' with the corresponding key words 'SIR' and 'FIR' below them appeared on the two sides of the screen in front of the participant. Listeners responded by pressing the corresponding button on a button box. Which prompt, key word, and button was on the left was counterbalanced across participants. Following the end of the stimulus, the participant had up to 1500 ms to respond. If feedback was given, it appeared in the center of the screen for 750 ms immediately after the participant responded or the response interval ended, whichever was earlier. Seven hundred fifty ms then elapsed before the next trial began.

Results

Figure 7 shows that at the shortest gate, 150 ms, when all the listeners heard was the initial fricative, the proportion of 'F' responses did not differ among the three following contexts. The proportion of 'F' responses also did not differ between those contexts at the next gate, 250 ms, but once listeners heard 350 ms of the stimuli, 'F' responses became noticeably more frequent in the *-ile* than the *-ime* or *-ide* contexts, an increase that got even larger at the next gate, 450 ms. At the longest gate, 705 ms, the preference for 'F' responses in the *-ile* context compared to the *-ime* and *-ide* contexts did not appear to be any greater than at the 450 ms gate.

The proportion of 'F' responses relative to 'S' responses served as the dependent variable in mixed effects logistic regression models in which the fixed effects were step along the /s/-/f/ continuum, gate, and lexical context. Lexical context was again treatment-coded such that responses in the two lexical bias context, *-ile* and *-ide*, were each compared to the no lexical bias context *-ime*. Step and gate were centered. Random effects of subject on the intercept and the slopes of the fixed effects and their interactions were included in the models, but not any correlations between the random effects. Table 4 lists the estimates for the fixed effects and their interactions from the model including all the

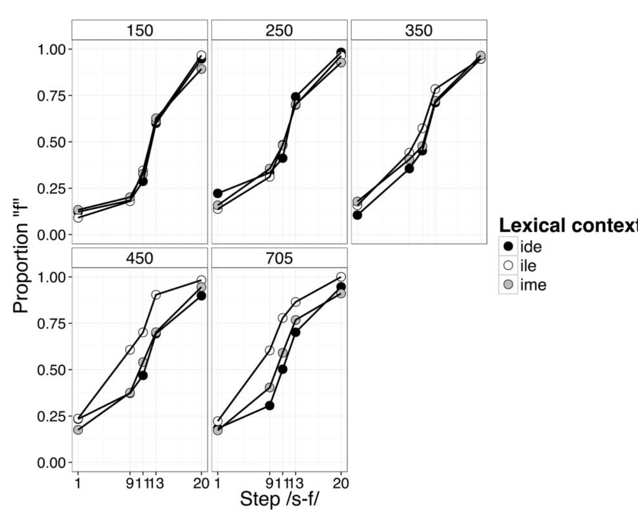

Figure 7. Experiment 1b 'F' response proportions as a function of gate, continuum step, and context.

Table 4
Parameter Estimates, Standard Errors, z Values, and p Values From A Mixed-Effects Logistic Regression Model of the Probability of an 'F' Response in Experiment 1b

Parameters	Estimate (β)	Std. Error	*z* Value	*p* Value
Intercept	.072	.330	.219	.827
Step	1.810	.259	6.976	<.001
-ide vs. -ime	−.209	.132	−1.576	.115
-ile vs. -ime	.601	.200	3.013	.003
Gate	.380	.085	4.455	<.001
Step × -ide vs. -ime	.138	.181	.762	.446
Step × -ile vs. -ime	.036	.135	.269	.788
Step × Gate	.025	.057	.428	.669
-ide vs. -ime × Gate	−.134	.093	−1.435	.151
-ile vs. -ime × Gate	.539	.153	3.524	<.001

two-way interactions between these fixed effects. The proportion of 'F' responses relative to 'S' responses increased significantly as the fricative became more /f/-like (positive step estimate) in the *-ile* context compared to the *-ime* context (positive *-ile* vs. *-ime*) and at longer Gates (positive gate estimate). There was a nonsignificant trend toward fewer 'F' responses in the *-ide* context compared to the *-ime* context (negative *-ide* vs. *-ime* estimate). The only significant interaction was between gate and the *-ile* versus *-ime* comparison. The positive parameter estimate shows that 'F' responses increased in the *-ile* context compared to the *-ime* context at longer gates.

To determine when listeners had heard enough of the context to activate their lexical knowledge, models including all two-way interactions were constructed for the following successive pairs of gates: 150 versus 250 ms, 250 versus 350 ms, and 350 versus 450 ms. In all three comparisons, 'F' responses increased significantly as the fricative became more /f/-like—that is, as step increased. In the model including 150 versus 250 ms gates, the proportion of 'F' responses did not differ between either the *-ile* or the *-ide* contexts and the *-ime* context, and context did not interact significantly with gate. In the model including 250 versus 350 ms gates, however, listeners responded 'F' significantly less often in the *-ide* than the *-ime* context ($\beta = -0.401$, $z = -1.983$, $p = .047$), and marginally more often in the *-ile* than the *-ime* context ($\beta = 0.380$, $z = 1.657$, $p = .098$). However, that positive bias was greater at the longer gate ($\beta = 0.727$, $z = 2.008$, $p = .045$). In the model with the 350 and 450 ms gates, listeners responded 'F' more often in the *-ile* than the *-ime* context ($\beta = 0.607$, $z = 2.703$, $p = .007$), and this preference was greater for the longer than the shorter gate ($\beta = 1.854$, $z = 3.727$, $p < .001$). For these gates, there was no significant difference in the proportion of 'F' responses in the *-ide* compared to the *-ime* context, either across the two gates ($\beta = -0.229$, $z = -1.404$, $p = .160$) or at one gate compared to the other ($\beta = 0.256$, $z = 0.692$, $p = .489$).

Discussion

This gating experiment showed that listeners heard at least a hint of the identity of the final consonant between 250 and 350 ms, enough that 'F' responses were modestly more frequent when they heard 350 ms of the *-ile* context compared to just 250 ms. However, 100 ms later, at the 450 ms gate, they had certainly heard enough of the consonant to strongly prefer an 'F' response in the *-ile* context. It's therefore not surprising that lexical knowledge would have induce more novel fixations to "F" shortly afterward in this context, in the 500–600 ms bin (Figures 5 and 6).

Experiment 2a

The results of Experiment 1a predict that if potentially lexically biasing information were presented at nearly the same time as the target sound, the two could affect the eye movement record nearly simultaneously. We tested this prediction in Experiment 2a. The target sounds in this experiment consisted of Steps 1, 4, 5, 6, 7, and 10 from a 10-step /ɛ/-/ʌ/ vowel continuum (i.e., from the vowel in the English word *leg* to the vowel in the word *lug*). The target was heard as the second segment in one of four syllabic contexts. In two of these contexts, *b-nk* (/b–ŋk/) and *d-nk* (/d–ŋk/), the /ʌ/ vowel produces a word (*bunk*, *dunk*) while the /ɛ/ vowel does not (**benk*, **denk*). In the other two contexts, *b-sh* (/b–ʃ/) and *d-sh* (/d–ʃ/), neither vowel produces a word (**besh*, **bush*, **desh*, **dush*). (While there is an English word *bush*, that word has a different vowel (/ʊ/) than that in *lug* (/ʌ/).) Thus, hearing an ambiguous vowel in the *b-nk* or *d-nk* context is expected to encourage an /ʌ/ response compared to the other two contexts. Critically, the following segment, which has the potential to activate lexical knowledge, becomes audible during or just after the vowel: The following /ŋ/ nasalizes that vowel, and both the following /ŋ/ and the following /ʃ/ influence the formant transitions at the end of the vowel distinctively. Again, we use a gating experiment with the same stimuli to test this assumption.

We also examined the effect of a second contextual manipulation to compare its timing to the effects we have already explored. This is the tendency to categorize an ambiguous sound as differing from a preceding sound (e.g., Holt, Lotto, & Kluender, 2000; Holt, 2006). This effect can either reflect compensation for coarticulation (e.g., Fowler, 2006) or auditory contrast (e.g., Lotto & Holt, 2006). Although the predicted compensation or contrast reflect very different conceptions of the listeners' mental processes, our results cannot distinguish these accounts. We hypothesized, nonetheless, that this effect in the eye movement record would be restricted to an early period, which would demonstrate that the methods used here can distinguish effects with different timing.

Spectral difference was manipulated by means of the initial segment. The /d/, as an alveolar consonant, has a higher second and third formant (F2 and F3) at the onset of the transition into the following vowel than its bilabial counterpart /b/. As a back vowel, /ʌ/ has a lower F2 and F3 than its front counterpart /ɛ/. This combination could result in a relative increase in /ʌ/ responses in the *d-nk* and *d-sh* contexts compared to the *b-nk* and *b-sh* contexts, which we hypothesize is independent from the lexical preference for /ʌ/ responses in the *d-nk* and *b-nk* contexts.

Method

Subjects. Twenty-nine different subjects from the same pool as Experiment 1 participated in Experiment 2.

Materials. As noted above, the target sounds consisted of Steps 1, 4, 5, 6, 7, and 10 from a 10-step /ɛ/-/ʌ/ continuum, and these sounds followed either /b/ or /d/. Table 5 lists the values of F2 and F3 in Hz at the onset of the transitions from /b/ and /d/ to

Table 5
Onset and Steady-State Values (in Hz) for F2 and F3 for Endpoint /bɛ-, bʌ-, dɛ-, and dʌ-/ Stimuli in Experiments 2a, b

Measurement	F2	F3
bɛ onset	1469	2461
bʌ onset	1100	2421
dɛ onset	1635	2744
dʌ onset	1542	2622
ɛ steady-state	1759	2650
ʌ steady-state	1131	2645

/ɛ/ and /ʌ/, and the steady-state values of each of these vowels 100 ms into the stimuli. Spectrograms of representative stimuli are displayed in Figure 8.

The noisy voiceless portions of these stimuli (the initial /b/ and /d/ bursts, the final /k/ burst, and the final /ʃ/) were taken from naturally produced stimuli and spliced before and after the voiced intervals; the final /k/ burst was separated from the end of the sonorant interval by a 50 ms-long interval of silence that simulated a voiceless stop closure. The vowel-nasal (VN) /ɛŋ/-/ʌŋ/ portions of the **benk-bunk* and **denk-dunk* continua, and the vowel (V) /ɛ/-/ʌ/ portions of the **besh-***bush* and **desh-***dush* continua were synthesized from parameter values measured in good exemplars of each syllable recorded by a male native speaker of American English. Formant frequencies and bandwidths were extracted from these exemplars, edited, and used as parameters in the Sensimetrics implementation of KLSYN88 (Klatt & Klatt, 1990). Intermediate steps in the continuum were constructed by linearly and incrementally interpolating between the endpoint values. The portions of the /ɛŋ/-/ʌŋ/ and /ɛ/-/ʌ/ intervals following the initial /b/ and /d/ consisted of formant transitions appropriate to bilabial and alveolar places of articulation (see Table 5 for onset and steady-state values of the two formants that determine both consonant place and vowel backness). Following these transition intervals, the members of the /ɛŋ/-/ʌŋ/ and /ɛ-ʌ/ continua did not differ between preceding /b/ and /d/ contexts. These continua also did not differ in their F0 contours. The durations of each interval are listed in Table 6, and the timing of these acoustic events can be seen in the spectrograms in Figure 8.

Each subject heard 84 experimental stimuli in each of the four contexts, for a total of 336 experimental trials. The 84 stimuli consisted of seven occurrences of Steps 1 and 10, 14 of Steps 4 and 7, and 21 of Steps 5 and 6. The 336 experimental trials were presented in a pseudorandomized order, with each successive set of 48 trials consisting of one occurrence of Steps 1 and 10, two of Steps 4 and 7, and three of Steps 5 and 6 in each of the four contexts. An initial training block of 48 trials, which were not included in the analysis, consisted of six occurrences of each of the endpoint stimuli in each of the four contexts, presented in randomized order with correct answer feedback. The initial training block was included in this experiment, unlike in Experiment 1a, to familiarize subjects with the less obvious mapping between target sounds and response letters (E and U).

Procedure. The subject's task was to respond whether /ɛ/ or /ʌ/ was heard on each trial by means of a mouse click on a visually presented uppercase letter E (displayed above the central fixation point) or U (displayed below the central point). The procedure was identical to Experiment 1a, with the exception that the central red circle changed to green on its own after 1 s, after which fixation on the central circle would trigger playing the auditory stimulus.

Results

The proportion of 'U' responses increased monotonically across the continuum, as shown in Figure 9. 'U' responses were also substantially more likely in the *b-nk* or *d-nk* contexts, where /ʌ/ makes a word, than in the *b-sh* or *d-sh* contexts where neither sound makes a word. This effect was largest at the midpoint steps, but also when the target was an unambiguous /ʌ/. Perhaps surprisingly, the difference at this endpoint is due to some 'E' responses to an unambiguous /ʌ/ sound in the contexts in which no lexical bias was expected. In addition, 'U' responses were more likely when the target sound followed /d/ than when it followed /b/.

As in Experiment 1a, mixed effects logistic regression models were used to evaluate these effects statistically. The fixed effects were continuum step (centered at 5.5), lexical context (coded with the neutral contexts taking the value −.5 and the lexically supportive contexts expected to display a 'U' preference taking the value .5), spectral context (coded with initial /b/ taking the value −.5 and initial /d/ taking the value .5), and the two- and three-way interactions between these factors. Random effects once again included random subject intercepts and random subject slopes for each of the main fixed effects. The results are shown in Table 7. There are significant effects of all three factors, as well as a significant interaction between continuum step and lexical context. As Figure 9 shows, this interaction reflects the larger effect of lexical context at the /ʌ/ end of the continuum.

Plots of fixation proportions on U over time as a function of continuum step and context are shown in Figures 10a and 10b. All three manipulations began to influence the location of fixations at about the same time, between 300 and 400 ms after stimulus onset.

We constructed the same two kinds of statistical models for this experiment as for Experiment 1a: The first examined effects of the experimental manipulations on the probability of fixating on U as opposed to E in each time window (i.e., the traditional analysis), and the second examined effects on the probability that a new fixation will fall on U as opposed to E (the novel analysis).

For the first analysis, the z scores corresponding to the main effects are shown in Figure 11a, and the z scores corresponding to the two-way interaction effects are shown in Figure 11b. (The three-way interaction never approached significance.) None of the three factors significantly affected fixation locations in the 200–300 ms bin, but all three factors had a significant influence in the 300–400 ms bin. The effect of continuum step persisted over the entire analysis period. Compared to continuum step, the effect of lexical context was somewhat shorter-lived but still relatively long lasting, while the effect of spectral context ended even sooner, by about 700 ms. The interaction between step and lexical context first reached significance in the 400–500 ms bin, and also persisted for a full second. There was also an apparent interaction between lexical and spectral context extending from the 1000 ms bin to the 1400 ms bin. Our confidence that this interaction is a meaningful effect is reduced, however, by its absence in the behavioral data and its failure to reach significance in the second analysis.

Figure 12 shows the corresponding z score plots from the analysis restricted to trials on which a new fixation on either U or

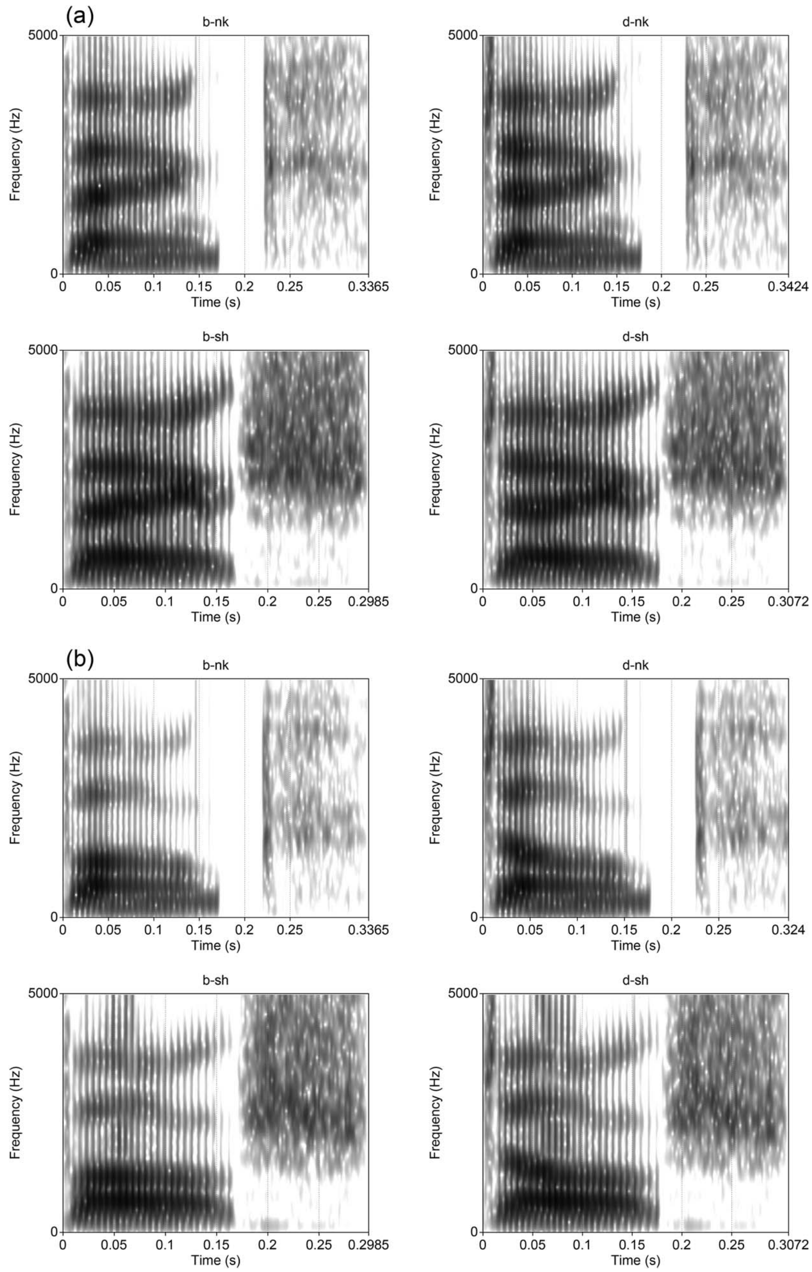

Figure 8. Spectrograms of (a) Step 1 and (b) Step 10 used in Experiments 2a, b: *b-nk* and *d-nk* (top), *b-sh* and *d-sh* (bottom). Labeled ticks mark the locations of the first five gates used in Experiment 2b.

E began in the time window of interest. These plots are truncated in the bin ending at 1300 ms, as after this point the statistical model would not converge. As for Experiment 1a, the results of the two analyses are broadly consistent, except that in this second analysis the effects once again do not persist as long. The ordering of the duration of the effects is the same (continuum step the longest, spectral context the shortest).

Figure 13a shows the proportion of new fixations in the 300–400 ms bin that fall on the letter U, by continuum step and context, and Figure 13b shows the proportion of new fixations in the 800–900 ms bin that fall on U, by continuum step and context. A fixation that begins in the 300–400 ms window is more likely to fall on U when the target sound is more /ʌ/-like, when the context is one in which /ʌ/ forms a word, and when the preceding sound is /d/ rather than /b/. By the 800–900 ms window, however, only continuum step continues to influence the location of new fixations. While both lexical context and the step x lexical context interaction continue to influence overall fixation proportions in this time window (see Figure 11), these figures reinforce the second of our statistical analyses (see Figure 12) in showing that these variables no longer influence the location of new fixations.

Table 6
Durations (in ms) of Various Intervals in Stimuli Used in Experiment 2a

Stimulus	Burst	VN/V	Stop Closure/Fricative	Burst
bVŋk	7.2	164	50	25
dVŋk	12.6	164	50	25
bVʃ	7.2	164	125	—
dVʃ	12.6	164	125	—

Discussion

The acoustics of the target sound, the potential lexical status of the stimulus, and the preceding sound all influenced the behavioral response at about the same time, 300–400 ms after stimulus onset. The stimuli used in this experiment, unlike in Experiment 1, were designed such that there was little or no delay between the target sound and the onset of acoustic information that was potentially relevant to a lexical bias. The experiment strongly confirms the suggestion from Experiment 1a that potentially lexically biasing information influences eye movements as rapidly as target sound acoustics themselves. The new factor manipulated in this experiment, spectral difference with the preceding sound, influences eye movements equally rapidly.

The effects differ in duration, however. The effect of spectral difference was very short-lived. This is arguably not surprising, as spectral difference between the target sound and the preceding sound would be expected to influence the perceived spectrum of the target's initial portion. More surprising, however, is that the target sound's lexical context actually influenced responses for a briefer period than its own acoustics. As Figure 13 shows, there is a clear effect of continuum step on the location of new eye fixations in the 800–900 ms window, but lexical context is no longer influencing the location of new fixations at that point.

The results of Experiment 1a are not actually inconsistent with the finding that lexical context effects are shorter-lived than the effects of target sound acoustics. In Experiment 1a, the two effects ended at about the same time, but the effect of lexical context began about 200 ms later than that of target sound acoustics, suggesting a shorter-lived effect overall. In Experiment 2a, where the two effects begin at the same time, the effect of lexical context ends earlier.

Experiment 2b

As in Experiment 1b, we tested our hypothesis about when listeners could first perceive acoustic properties that would inform them that the vowel was followed by /ŋk/ or /ʃ/ by running a gating study with a new group of participants.

Methods

Subjects. Participants were 20 different people from the same pool as Experiment 1b.

Materials. The endpoints, Steps 1 and 10, and the two most frequently presented stimuli from Experiment 2a, Steps 5 and 6, were used in this experiment. The stimuli were gated the same way as they were in Experiment 1b, except that the gates were more closely spaced, at 50, 100, 150, 200, 250, and 405 ms, because the stimuli were shorter (see Figure 8). The longest gate again extended beyond the end of the stimuli. The tick marks on the lower borders on the spectrograms of the representative stimuli in Figure 8 show where each of the gates was applied. The shortest gate at 50 ms consisted of just the initial stop burst and the transition to the vowel, the second at 100 ms extended a little more than halfway through the vowel, and the third at 150 ms extended to just before the end of the vowel. The fourth gate at 200 ms occurred either during the stop closure for the final /k/ in the *bVnk* and *dVnk* stimuli or early in the fricative for the *bVsh* and *dVsh* stimuli, and the fifth at 250 ms fell after the noise burst at the release of the final /k/ or little more than halfway through the final fricative /ʃ/. The longest gate at 405 ms, which fell after the end of the stimulus, is not shown.

Procedure. Listeners were first presented with 48 training trials in which the endpoints (Steps 1 and 10) in each of the two initial and two final consonant contexts at each of the six gates were presented once with correct answer feedback. Training was followed by three blocks of test trials in which Steps 1, 5, 6, and 10 in each initial and final context at each gate were presented in a ratio of 1:5:5:1, for a total 288 test trials per block and 864 test trials per listener. Correct answer feedback was still given to the endpoint stimuli during the test blocks. In both training and test blocks, stimuli were presented in a different random order to each participant. Participants took brief self-timed breaks one-third and two-thirds of the way through each block (every 96 trials), and longer breaks between blocks.

At the same time as the stimulus began to play on each trial, the response prompts "E" and "U" with the corresponding key words "LEG" and "LUG" below them appeared on the two sides of the screen in front of the participant. Listeners responded by pressing the corresponding button on a button box. Which prompt, key word, and button was on the left was counterbalanced across participants. Following the end of the stimulus, the participant had up to 1500 ms to respond. If feedback was given, it appeared in the center of the screen for 750 ms immediately after the participant responded or the response interval ended, whichever was earlier. Seven hundred fifty ms then elapsed before the next trial began.

Results

Figure 14 is a plot of the proportion of 'U' responses averaged across participants. Each panel represents responses to a single

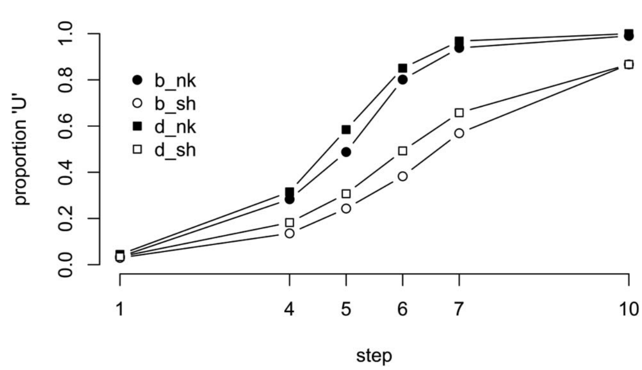

Figure 9. Experiment 2a 'U' response proportions as a function of continuum step and context.

Table 7
Parameter Estimates, Standard Errors, z Values, and p Values From a Mixed-Effects Logistic Regression Model of the Probability of a ‘U’ Response in Experiment 2a

Parameters	Estimate (β)	Std. Error	*z* Value	*p* Value
Intercept	.164	.242	.68	.496
Step	1.486	.095	15.56	<.001
Lexical Context	2.288	.568	4.03	<.001
Spectral Context	.531	.154	3.46	<.001
Step × Lexical Context	.719	.065	11.13	<.001
Step × Spectral Context	.027	.057	.47	.638
Lexical Context × Spectral Context	−.132	.136	−.97	.332
Step × Lexical Context × Spectral Context	.031	.116	.27	.785

gate. At the 100 ms gate, there is a hint that the consonant(s) following the vowel influences responses: for both initial consonants, listeners responded ‘U’ more often when /ŋk/ followed than when /ʃ/ did. By the gate at 150 ms, the influence of the following consonants is clear and remains robust for all longer gates. At the shorter gates, and especially at 50 ms, participants responded ‘U’ more often after /b/ than /d/. This finding is unexpected because Experiment 2a showed that participants responded ‘U’ more often after /d/ than /b/. We discuss this discrepancy in the current section.

The proportion of ‘U’ relative to ‘E’ responses served as the dependent variable in mixed effects logistic regression models in which the fixed effects were steps along the /ɛ/-/ʌ/ continuum (increasing from /ɛ/ to /ʌ/) centered at 5.5, spectral context (= initial consonant: /b/ = −0.5, /d/ = 0.5), lexical context (= final consonant(s): /ʃ/ = −0.5, /ŋk/ = 0.5), and gate. Random effects of subject were included on the intercept and slopes of the fixed effects and any interactions between them; however, correlations between random effects were omitted to ensure that the models converged.

Estimates from a model including all possible three-way interactions are listed in Table 8. Listeners responded ‘U’ significantly more often as step increased (positive step estimate) and before following /ŋk/ compared to following /ʃ/ (positive lexical context estimate), but less often after preceding /d/ than /b/ (negative spectral context estimate). Gate did not affect the likelihood of a

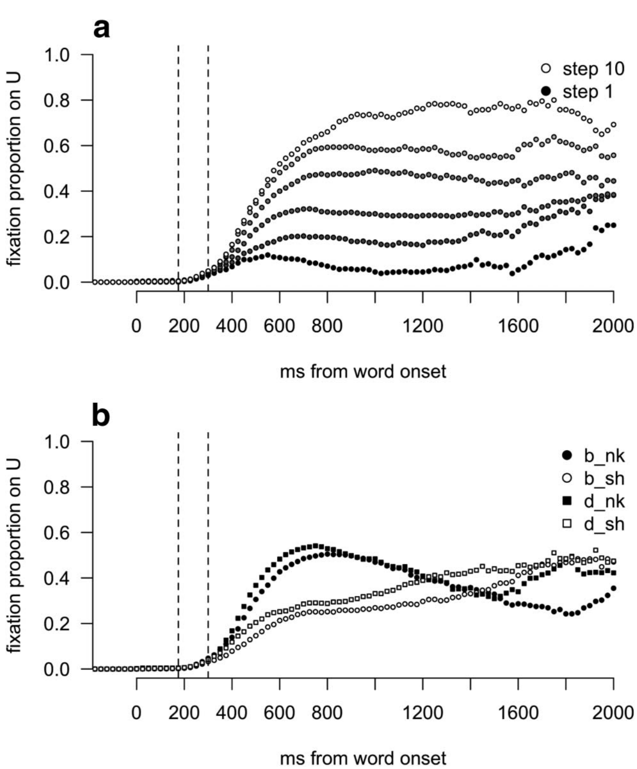

Figure 10. Experiment 2a fixation proportion on U over time as a function of (a) continuum step and (b) context. Endpoint steps are excluded from the latter figure. Vertical lines at 175 ms and 300 ms represent the approximate offsets of the voiced and following voiceless portions of the stimuli.

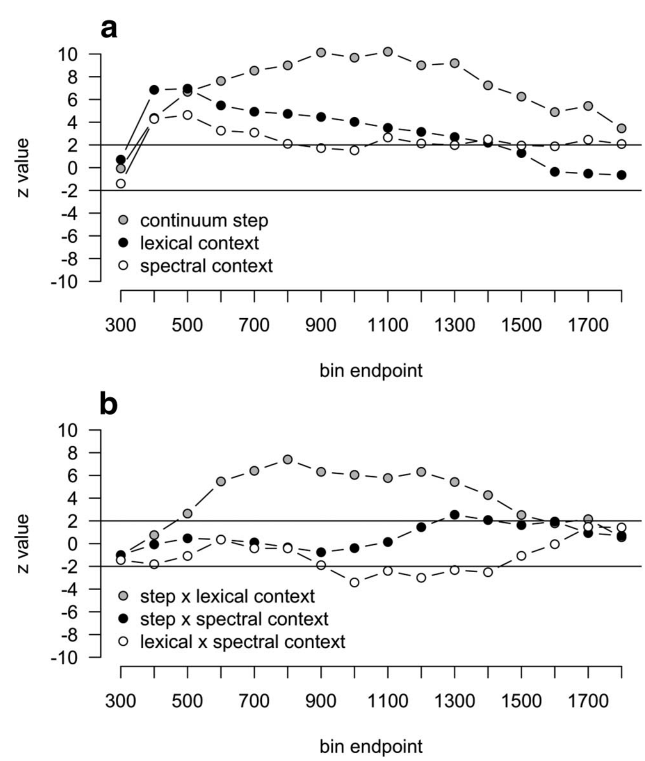

Figure 11. Experiment 2a *z* scores from a mixed effects model of eye movement data for (a) the three main effects and (b) the three two-way interaction effects.

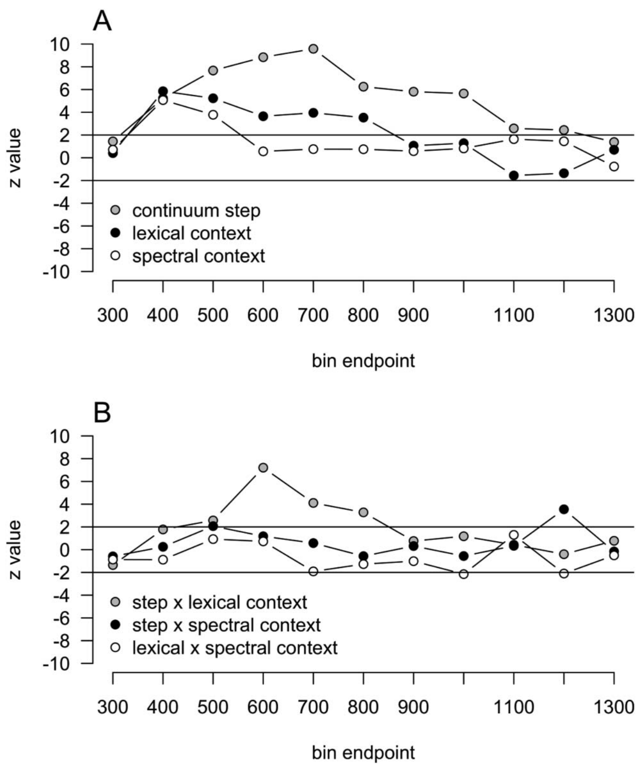

Figure 12. Experiment 2a z scores from a mixed effects model of eye movement data for (a) the three main effects and (b) the three two-way interactions, from the alternate analysis described in the text.

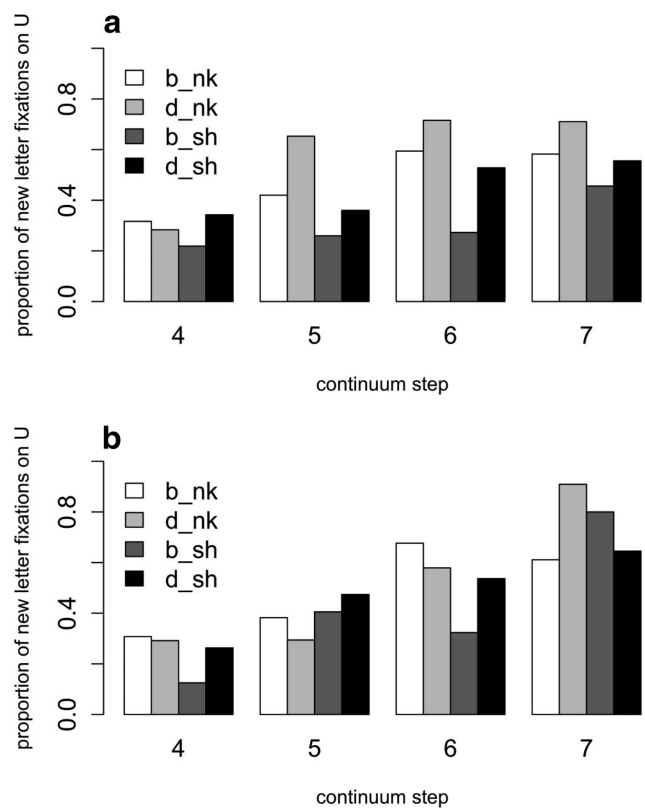

Figure 13. Experiment 2a proportion of new fixations on a letter in the 300–400 ms bin (a) and in the 800–900 ms bin (b) that fall on U, as a function of context and step.

'U' response by itself. Three two-way interactions were significant: participants responded 'U' more often as step increased for longer gates (positive step by gate interaction), more often before following /ŋk/ than following /ʃ/ for longer gates (positive lexical context by gate interaction), and more often after preceding /d/ than preceding /b/ for longer gates (positive spectral context by gate interaction). This last interaction reflects the shrinkage of the very large difference in 'U' responses after preceding /b/ compared to preceding /d/ at gates longer than 50 ms and not an actual preference for 'U' responses after preceding /d/.

To further narrow down the interval in which following /ŋk/ activated the lexical items *bunk* or *dunk* and created the lexical bias in favor of 'U' responses, we constructed similar models for two adjacent pairs of gates, first for gates at 50 and 100 ms and second for gates at 100 and 150 ms. Clear evidence for lexical activation was obtained in the comparison of the 100 ms gate with the 50 ms gate: Participants responded 'U' significantly more often before following /ŋk/ than following /ʃ/ (lexical context: $\beta = 0.650$, $z = 4.314$, $p < .001$) and more so for the longer 100 ms gate than the shorter 50 ms one (lexical context by gate: $\beta = 0.524$, $z = 3.751$, $p < .001$). Figure 14 shows that lexical preference strengthens considerably between the 100 and 150 ms gates. After the longer 150 ms gate compared to the shorter 100 ms gate, the relative proportion of 'U' responses also increased significantly before following /ŋk/ compared to following /ʃ/ (lexical context by gate interaction: $\beta = 0.333$, $z = 2.273$, $p = .023$). In both models, 'U' responses increased after preceding /d/ compared to preceding /b/ in the later gate (spectral context by gate interaction at 50 vs. 100 ms: $\beta = 1.025$, $z = 6.147$, $p < .001$; at 100 vs. 150 ms: $\beta = 0.501$, $z = 2.919$, $p = .004$), which reflects the progressive shrinkage of the very strong preference for 'U' responses after preceding /b/ compared to after preceding /d/ at the earliest gate.

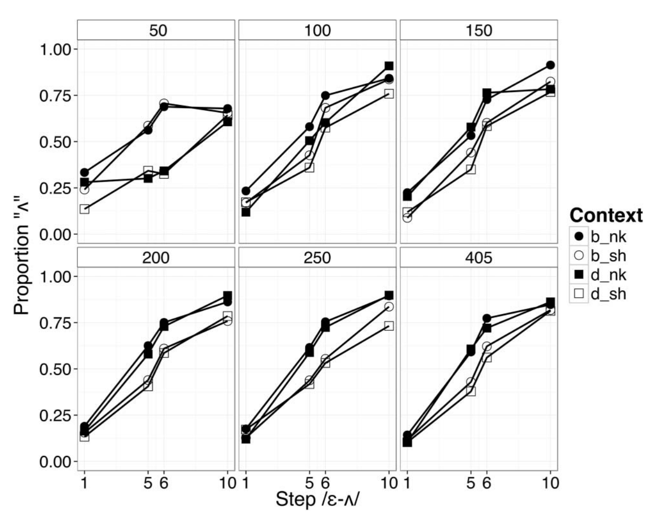

Figure 14. Experiment 2b 'U' response proportions as a function of gate, continuum step, and context.

Table 8
Parameter Estimates (β), Standard Errors, z Values, and p Values From a Mixed-Effects Logistic Regression Model of the Probability of a 'U' Response in Experiment 2b

Parameters	Estimate (β)	Std. Error	*z* Value	*p* Value
Intercept	.297	.205	1.449	.147
Step	1.189	.168	7.057	<.001
Lexical Context	.372	.166	2.238	.025
Spectral Context	−.211	.060	−3.550	<.001
Gate	.180	.137	1.312	.190
Step × Lexical Context	.015	.070	.207	.836
Step × Spectral Context	−.028	.033	−.856	.392
Step × Gate	.380	.062	6.080	<.001
Lexical Context × Spectral Context	.020	.024	.839	.401
Lexical Context × Gate	.210	.107	1.970	.049
Spectral Context × Gate	.151	.034	4.413	<.001
Step × Lexical Context × Spectral Context	.012	.024	.510	.610
Step × Lexical Context × Gate	.087	.063	1.379	.168
Step × Spectral Context × Gate	.003	.025	.112	.911
Lexical Context × Spectral Context × Gate	.005	.021	.222	.825

Discussion

The results of this gating experiment show that listeners could indeed have heard acoustic properties of the final consonants in these stimuli quite early during the vowel, apparently by no more than 100–150 ms after the stimulus began. Specifically, they could have heard that the vowel was nasalized preceding /ŋk/, which could have activated either *bunk* or *dunk* and thereby encouraged early looks toward 'U' in that context. Hearing an unnasalized vowel would have been evidence that the final consonant was instead /ʃ/, and that percept would not have activated any lexical item nor encouraged one response more than another. This access to acoustic evidence that would distinguish stimuli that activate a lexical item from those that would not is early enough to significantly increase new fixations to 'U' in the *b-nk* and *d-nk* contexts in the 400–500 ms bin (Figures 12 and 13).

Fifty ms of the stimulus was sufficient for listeners to be sensitive to the acoustic manipulation of the vowel, but the effect of the following consonant did not appear until the 100 ms gate. This apparently contrasts with the results of Experiment 2a, where the onset of these effects did not differ in eye movement record. This is probably due to the limited ability of the eye movement experiment to detect the subtle sequencing of effects that the gating experiment was explicitly designed to test. Close inspection of the fixation proportion curves in Figure 10 does suggest, however, that continuum step may affect eye movements very slightly earlier than lexical context.

Another aspect of these gating results is puzzling: Listeners did not respond 'U' more often after initial /d/ than initial /b/ as they did in Experiment 2a, but instead showed the opposite preference at early gates. The strong preference for 'U' responses after /b/ at the shortest gate, 50 ms, is explicable as a misperception of the lower F2 and F3 frequencies in the transition from /b/ to the vowel as evidence that the vowel was back. That preference shrinks at longer gates as listeners separate the acoustic evidence in the transitions perceptually from that in the following steady-state. Nonetheless, they never responded 'U' more often after /d/ than /b/ even at the longest gate, when they heard the entire stimulus. Perhaps the presence of the 50 ms gate and the strong preference for 'U' responses to it when the initial consonant was /b/ inhibited listeners from perceiving /ʌ/ more often after /d/ as they did in Experiment 2a when they heard the entire stimulus on every trial. Testing this speculation must wait for another study.

The difference between when lexical knowledge biased responses in the gating experiment and in the corresponding eye tracking experiment appears to be smaller in Experiment 1 than Experiment 2. The gating results from Experiment 1b (see Figure 7) show that acoustic evidence that would distinguish final /l/ from /m/ first became audible at the 350 ms gate and robustly influenced responses at the 450 ms gate. That evidence first influenced fixation proportions significantly for the *-ile* context relative to the *-ime* context in the 600–700 ms bin (Figure 5A), or 150–250 ms later. The gating results from Experiment 2b (see Figure 14) show that acoustic evidence that would distinguish final /ŋk/ from /ʃ/ first became audible at the 100 ms gate and robustly influenced responses at the 150 ms gate. That evidence first affected fixation proportions in the 400–500 ms bin, or 250–300 ms later. It appears to take roughly 100 ms longer for that evidence to affect fixation proportions relative to when it biases responses to the gated stimuli in Experiment 2 than in Experiment 1.

There are many differences between the two experiments that may play a role in explaining this difference. For example, the target sound is a vowel in Experiment 2 but a consonant in Experiment 1; this sound is medial in the syllable in Experiment 2 but initial in Experiment 1; and the lexically supportive consonant follows the target immediately in Experiment 2 but is separated from it by a prolonged diphthong in Experiment 1. None of these suggestions are actually potential explanations, however, until they are provided with a mechanism that would delay the influence of lexically supportive acoustic evidence on fixation proportions relative to its influence on categorizing gated stimuli under conditions like those in Experiment 2 compared to conditions like those in Experiment 1. Running the additional experiments needed to discover that mechanism is beyond the scope of the current article.

General Discussion

This article reports two experiments that used eye tracking to measure the timing of the Ganong effect, that is, the preference to identify an acoustically ambiguous stimulus as the category that forms a word with its context. In the first of these experiments, listeners identified the members of an /s/-/f/ continuum more often as the category /f/ in the context *-ile*, where /f/ but not /s/ forms a word, than in the neutral context *-ime*, where neither /f/ nor /s/ forms a word. Listeners also identified the members of this continuum slightly less often as /f/ in the context *-ide*, where /s/ but not /f/ makes a word, than in the neutral context. In the second experiment, they identified the members of an /ɛ/-/ʌ/ continuum more often as /ʌ/ in the contexts *b-nk* and *d-nk*, where /ʌ/ makes a word and /ɛ/ does not, than in the contexts *b-sh* and *d-sh*, where neither category makes a word. Manipulation of the initial consonant in Experiment 2 created an additional bias toward an /ʌ/ response after /d/, versus a bias toward an /ɛ/ response after /b/.

The two experiments differed in how soon after the target sound acoustic information became audible that could activate lexical knowledge. In Experiment 1a, this information was in a word-final consonant following a relatively long nucleus, while in Experiment 2a, this information was contained in the transitions to the following consonant at the end of the target sound itself. The results of the corresponding gating experiments, 1b and 2b, showed that acoustic evidence that could activate words was heard as early as 350 ms into the stimuli in Experiment 1a and as early as 100 ms into the stimuli in Experiment 2a.

Unlike RT data from mouse clicks or button presses, eye movements are fast and are planned and executed while the stimulus is still unfolding. Eye movements were analyzed with two methods. The first method was a traditional one in which the dependent variable was the relative proportion of all eye fixations in each 100 ms bin that fell on one of the two letters (F in the first experiment, U in the second). The second method again divided the eye movement data into 100 ms bins, but included only those trials on which there was a new eye fixation on a letter. Because fixations that began in any previous bin and continued into the current one are included in the first method, the proportions in each bin capture the cumulative history of the participants' eye movements. The second method instead captures current effects of stimulus characteristics on new eye movements to the target letter, launched from the other letter or elsewhere.

In Experiment 1a, both methods showed that the target sound's acoustics (step) influenced fixations earlier than the effect of the listeners' lexical knowledge, and neither showed an interaction of the two effects (Figures 4–6). However, lexical knowledge influenced fixations very quickly relative to when lexically supportive acoustic information first became audible in the stimulus, late in the vowel preceding the final consonant. In Experiment 2a, both methods showed that the target sound's acoustics, the listeners' lexical knowledge, and the spectral difference with the preceding consonant all began to influence eye movements during the 300–400 ms bin. Both methods also showed that the acoustic effect lasted longest and the contrast effect least long, and that an interaction between the acoustic and lexical effects emerged later than the individual effects did, in the 500–600 ms bin. Otherwise, the first method shows, unsurprisingly, all effects lasting longer than the second, given that the second method doesn't count fixations that persist from previous time bins.

Arguments can be made for both methods, as they reveal different aspects of eye movement behavior. However, it is also worth noting that the novel method most closely reflects the 'linking hypothesis' suggested by Allopenna, Magnuson, and Tanenhaus (1998), relating eye movements in the visual world paradigm to activation of linguistic and cognitive representations. In the context of an experiment in which the visual stimuli were pictures corresponding to spoken words, Allopenna et al. proposed that:

> "the probability of initiating an eye movement to fixate on a target object *o* at time *t* is a direct function of the probability that *o* is the target given the speech input and where the probability of fixating *o* is determined by the activation of its lexical entry relative to the activations of the other potential targets" (1998, p. 424).

They also write that the dependent measure of interest is "the probability of directing attention to an object in space along with a concomitant eye movement" (1998, p. 424). This emphasis on the movement of eyes, rather than their location, is entirely appropriate, though it does not correspond to the method actually used by Allopenna et al., or most subsequent researchers using the visual world paradigm. It is our novel method that analyzes effects on eye movements, as opposed to effects on the eyes' current gaze location. Indeed, effects on eye movements would be analyzed most directly by measuring when target-directed saccades are *launched*, rather than when target fixations begin. Given that most saccades take much less than 100 ms, however, and that their duration is unlikely to differ across experimental conditions, such an analysis is likely to replicate the analysis we report here.

The effect of lexical knowledge appeared quite early in both experiments, relative to the point at which the acoustic properties of the potentially lexically biasing segment became audible, as established by the gating experiments (1b and 2b). Indeed, the effect of lexical knowledge on eye movements appears to be about as fast as the effect of the target sound's acoustics. This finding argues against treating the Ganong effect as the product of a temporally extended feedback mechanism, as TRACE does (McClelland & Elman, 1986). Feedback takes substantial time to influence phoneme activation; in McClelland and Elman's (1986) simulation, the delay was equivalent to at least the duration of two segments. While changes to specific model parameters could reduce this delay, any account of the Ganong effect that requires spreading activation from phonemes to words, and back to phonemes, must posit a delay of some duration. On the other hand, Merge (Norris et al., 2000; McQueen et al., 2003) predicts a relatively more immediate effect. While lexical activation still takes time, any delay in lexical activation's influence on phonemic categorization need not include the time taken to feed back to the phonemic level.

The TRACE model also predicts that the lexical effect should gradually increase in strength over the course of a trial. The present experiments show that this is not the case. Our novel analysis of the eye movement data, which allows assessment of independent effects in each time window, shows that the influence of lexical context is, if anything, shorter-lived than the influence of target sound acoustics itself. Figures 12 and 13 show that in Experiment

2a, there is clearly a period after which the lexical context no longer has any effect, but the acoustics of the target sound still do.

These results appear to be consistent with the results of McQueen et al. (2003), who found that the Ganong effect was largest for fast responses even when the target was syllable-initial. The results are not consistent with studies (Fox, 1984; Miller & Dexter, 1988; Pitt & Samuel, 1993) that found a larger Ganong effect for slow responses when the target was syllable-initial. Whatever causes the discrepancy between these RT studies, the present eye movement experiments make clear that lexical information has a very rapid but decreasing effect over the course of a trial. The gating studies also show that acoustic information that could activate words can be heard and used very early in the stimuli.

Recent work by Gow, Segawa, Ahlfors, and Lin (2008) appears to challenge an autonomous strictly feed-forward model in which prelexical processing is uninfluenced by lexical feedback. Gow et al. applied a Granger-causality analysis to magnetoencephalography (MEG) and electroencephalography (EEG) data to assess causal relations between brain regions that are active during auditory processing (principally the left posterior superior temporal gyrus, L pSTG), during lexical activation (principally the left supramarginal gyrus, L SMG), and during explicit categorization of speech sounds (principally the left inferior frontal gyrus, L IFG) in a Ganong-type experiment. The target sound was a member of an /s/-/ʃ/ fricative continuum followed by contexts where one of these categories made a word and the other did not, *-andal* (*sandal*, **shandal*) versus *-ampoo* (**sampoo*, *shampoo*). When the fricative was the intermediate step along the /s/-/ʃ/ continuum that elicited the largest lexical effect on categorization (referred to as the "$\text{Ganong}_{\text{Max}}$" stimulus by Gow et al.), activity in the L SMG and the L IFG were both causally influenced by L pSTG activity during an early period from 80–280 ms after stimulus onset. During a later period, from 280–480 ms, the first causal relationship reversed, with L pSTG activity being influenced by L SMG activity. The fricative ended 120 ms after stimulus onset and the acoustic evidence that distinguished the two contexts lexically became audible 220 ms after stimulus onset, so this later interval includes the portion of the signal that activates one word or the other. Finally, activity in the L IFG in response to this stimulus was influenced by L pSTG activity during the latest period, from 480–750 ms after stimulus onset.

Gow et al. interpret the sequential timing of these causal relations during the three periods as arguing that the Ganong effect cannot be produced by a strictly feed-forward process. Instead, prelexical auditory processing in L pSTG first feeds forward to lexical and phonemic processing in L SMG and L IFG during the earliest period, which in turn feeds back to auditory processing in L pSTG during the middle period, which finally feeds forward to a decision about category identity in L IFG in the latest period. This sequence of events corresponds closely to the mechanism proposed by TRACE. Moreover, since the Granger causality analysis links activity in L pSTG but not activity in L SMG to L IFG activity in the latest period, it appears that lexical activation in L SMG is not merged with perceptual processing in L pSTG in arriving at a phoneme decision in L IFG. The timing of the Ganong effect revealed in the present experiments, by the behavioral method with the best available temporal resolution (i.e., eye movements) thus appears to conflict with the inference about the mechanism that is supported by Granger causality analysis.

This review of Gow et al.'s results and their interpretation has so far hewed closely to their own account; however, the results of their Granger causality analyses displayed in Figure 2 in their supplemental materials are more complex in ways that suggest that their interpretation is not straightforward. We first discuss causal relationships among activity in different brain areas during the three time periods in response to the $\text{Ganong}_{\text{Max}}$ stimulus, and then turn to comparing those causal relationships with those observed in response to the unambiguous word and nonword stimuli. Because a great many causal relationships were found in each period and to each of the three kinds of stimuli, only the most robust relationships between left hemisphere brain areas are discussed.

In response to the $\text{Ganong}_{\text{Max}}$ stimulus during the early 80–280 ms period, activity in L pSTG is causally related to activity in the left *anterior* superior temporal gyrus (L aSTG), and L IFG activity is causally related during this period to activity in both L pSTG and L aSTG. A notable absence during this period is any causal relationship between L SMG activity and L pSTG activity. Instead left angular gyrus activity (L AG) is causally related to L pSTG activity, and L SMG activity is more weakly causally related to L AG activity. During the middle 280–480 ms period, activity in L pSTG is not only causally related to L SMG and L IFG activity, but continues to be causally related to L aSTG activity. Activity in L IFG also continues to be causally related to L aSTG activity during this period, and a causal relationship emerges of L SMG to L aSTG activity. Activity in L AG is also causally related to activity in L pSTG, L IFG, and L SMG. Finally, in the late 480–780 ms period, only L IFG activity is robustly causally related to L pSTG activity. This large number of causal relationships and their appearances and disappearances in these three periods suggests that the sequence described earlier of causal relationships and the processes they implied—auditory analysis in the L pSTG, to lexical activation in L SMG, to feedback to continued auditory processing in L pSTG, and at last to phoneme categorization in L IFG—is at best incomplete. They also suggest that Gow et al.'s results cannot be used to distinguish between bottom-up and top-down causal relationships nor to ascribe particular mental processes either to activity in particular brain areas or to causal relationships between activity in one brain area and another; see Norris, McQueen, and Cutler (2016) for a similar critique of Gow et al.'s interpretation.

The interpretative difficulty is magnified by the very different causal relationships observed during the three time periods in response to the $\text{Ganong}_{\text{Max}}$ stimulus and to the unambiguous word and nonword stimuli (*sandal*, *shampoo*, **shandal*, **sampoo*). While it's not surprising that causal relationships between activity in brain areas would differ between responses to unambiguous words and unambiguous nonwords, it is surprising that these relationships differ between responses to the $\text{Ganong}_{\text{Max}}$ stimulus and both unambiguous words and nonwords. On the one hand, causal relationships reflecting the responses of brain areas to the $\text{Ganong}_{\text{Max}}$ stimulus might be expected to resemble those of unambiguous nonwords during the early period because the acoustics of the ambiguous fricative in the $\text{Ganong}_{\text{Max}}$ stimulus differs from the one expected in both the *-andal* and *-ampoo* contexts. On the other hand, it might be expected that the causal relationships would resemble those of unambiguous words during the middle and perhaps also the late period because the lexicon has biased the categorization of this fricative toward the category that makes a word in its context. Neither expectation is met, however.

We argue, therefore, that Gow et al.'s results neither rule out a strictly feed-forward model nor undermine the interpretation of our results as supporting such a model.

Conclusion

The present experiments show that acoustic information that can activate a word influences eye movements immediately in a phonemic categorization task, and that this influence also diminishes quite rapidly; indeed, perhaps more rapidly than the influence of target sound acoustics. This finding contradicts the predictions of an account of the Ganong effect according to which it depends on gradually developing feedback between phonemic and lexical levels (McClelland & Elman, 1986). In addition, the present article offers an alternative method for analyzing eye movements in the visual world paradigm, which focuses on the timing with which experimental manipulations influence the probability of new eye fixations landing on a given stimulus. Finally, the gating experiments show that listeners hear and use acoustic evidence that could activate words as soon as it occurs in the stimuli.

References

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439. http://dx.doi.org/10.1006/jmla.1997.2558

Altmann, G. T. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137,* 190–200. http://dx.doi.org/10.1016/j.actpsy.2010.09.009

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1. 1–7, http://CRAN.R-project.org/package=lme4

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers, 41,* 977–990. http://dx.doi.org/10.3758/BRM.41.4.977

Connine, C. M., & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 291–299. http://dx.doi.org/10.1037/0096-1523.13.2.291

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19,* 141–177.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68,* 161–177. http://dx.doi.org/10.3758/BF03193666

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 526–540. http://dx.doi.org/10.1037/0096-1523.10.4.526

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125. http://dx.doi.org/10.1037/0096-1523.6.1.110

Gow, D. W., Jr., 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,* 614–623. http://dx.doi.org/10.1016/j.neuroimage.2008.07.027

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America, 121,* 2899–2914. http://dx.doi.org/10.1121/1.2534656

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 592–608. http://dx.doi.org/10.1037/0096-1523.24.2.592

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America, 120,* 2801–2817. http://dx.doi.org/10.1121/1.2354071

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108,* 710–722. http://dx.doi.org/10.1121/1.429604

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820–857.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics, 68,* 178–183. http://dx.doi.org/10.3758/BF03193667

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. http://dx.doi.org/10.1207/s15516709cog2702_6

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753–771. http://dx.doi.org/10.1037/0096-1523.9.5.753

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

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10,* 363–369. http://dx.doi.org/10.1016/j.tics.2006.06.007

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1609–1631. http://dx.doi.org/10.1037/a0011747

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15,* 1064–1071. http://dx.doi.org/10.3758/PBR.15.6.1064

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33–B42. http://dx.doi.org/10.1016/S0010-0277(02)00157-9

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,* 433–443. http://dx.doi.org/10.1037/0096-1523.17.2.433

McQueen, J. M., Cutler, A., & Norris, D. (2003). Flow of information in the spoken word recognition system. *Speech Communication, 41,* 257–270. http://dx.doi.org/10.1016/S0167-6393(02)00108-5

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology, 60,* 661–671. http://dx.doi.org/10.1080/17470210601183890

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology:*

Human Perception and Performance, 14, 369–378. http://dx.doi.org/10.1037/0096-1523.14.3.369

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465. http://dx.doi.org/10.3758/BF03213823

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505–512. http://dx.doi.org/10.3758/BF03208147

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language, 69,* 527–545. http://dx.doi.org/10.1016/j.jml.2013.07.002

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition, 84,* 55–71. http://dx.doi.org/10.1016/S0010-0277(02)00014-8

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299–325. http://dx.doi.org/10.1017/S0140525X00003241

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition, and Neuroscience, 31,* 4–18. http://dx.doi.org/10.1080/23273798.2015.1081703

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290. http://dx.doi.org/10.3758/BF03213946

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39,* 347–370. http://dx.doi.org/10.1006/jmla.1998.2571

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,* 699–725. http://dx.doi.org/10.1037/0096-1523.19.4.699

R Core Team. (2014). R: A Language and Environment for Statistical Computing [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rysling, A., Kingston, J., Staub, A., Cohen, A., & Starns, J. (2015). Early Ganong effects. *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow, Scotland, August 10–14, 2015, 5 pp.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89. http://dx.doi.org/10.1016/S0010-0277(03)00139-2

Smits, R., Warner, N., McQueen, J. M., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *The Journal of the Acoustical Society of America, 113,* 563–574. http://dx.doi.org/10.1121/1.1525287

Staub, A., Abbott, M., & Bogartz, R. S. (2012). Linguistically guided anticipatory eye movements in scene viewing. *Visual Cognition, 20,* 922–946. http://dx.doi.org/10.1080/13506285.2012.715599

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Retrieved from http://www.iphod.com

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9,* 325–329. http://dx.doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408. http://dx.doi.org/10.1006/jmla.1998.2618

Received December 29, 2014
Revision received May 6, 2016
Accepted May 11, 2016 ■