



Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition

Bob McMurray^{a,*}, Michael K. Tanenhaus^b, Richard N. Aslin^b

^a Department of Psychology, E11 SSH, University of Iowa, Iowa City, IA 52240, USA

^b Department of Brain and Cognitive Sciences, University of Rochester, USA

ARTICLE INFO

Article history:

Received 5 December 2007
revision received 4 July 2008
Available online 19 September 2008

Keywords:

Spoken word recognition
Speech perception
Gradiency
Eye movements
Lexical ambiguity
Attractor dynamics
Phoneme categorization
TRACE model of word recognition
Mismatch

ABSTRACT

Spoken word recognition shows gradient sensitivity to within-category voice onset time (VOT), as predicted by several current models of spoken word recognition, including TRACE (McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86). It remains unclear, however, whether this sensitivity is short-lived or whether it persists over multiple syllables. VOT continua were synthesized for pairs of words like *barricade* and *parakeet*, which differ in the voicing of their initial phoneme, but otherwise overlap for at least four phonemes, creating an opportunity for “lexical garden-paths” when listeners encounter the phonemic information consistent with only one member of the pair. Simulations established that phoneme-level inhibition in TRACE eliminates sensitivity to VOT too rapidly to influence recovery. However, in two Visual World experiments, look-contingent and response-contingent analyses demonstrated effects of word initial VOT on lexical garden-path recovery. These results are inconsistent with inhibition at the phoneme level and support models of spoken word recognition in which sub-phonetic detail is preserved throughout the processing system.

© 2008 Elsevier Inc. All rights reserved.

Speech perception has been classically framed as a problem of overcoming variability in the signal to recover underlying linguistic categories, such as features, phonemes and words. This variability arises because the signal unfolds as a transient series of acoustic events created by partially overlapping articulatory gestures. These gestures are conditioned by the segment currently being uttered and by the properties of preceding and upcoming segments. Such processes impose significant variability on the signal. Even multiple utterances of the same word produced by a single speaker in a consistent context show significant variability (Newman, Clouse, & Burnham, 2001). Therefore, different tokens of the same sound and same word are likely to vary along a number of different dimensions.

While acoustic cues are variable, listeners must ultimately make a more or less discrete decision about word identity. Although early work suggested that within-cate-

gory variation was lost during categorization, especially for consonants (Ferrero, Pelamatti, & Vaggel, 1982; Kopp, 1969; Larkey, Wald, & Strange, 1978; Liberman, Harris, Hoffman & Griffith, 1957; Liberman, Harris, Kinney & Lane, 1961), more recent research demonstrates that spoken word recognition is exquisitely sensitive to sub-phonetic variation. For example, listeners use segmental duration to help distinguish between a monosyllabic word, such as *ham*, and a potential carrier word, such as *hamster* (Davis, Gaskell, & Marslen-Wilson, 2002; Gow & Gordon, 1995; Salverda, Dahan, & McQueen, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov & McDonough, 2007). Misleading coarticulatory information in vowels delays recognition, especially when the misleading information is temporarily consistent with another lexical candidate (Dahan, Magnuson, Tanenhaus, & Hogan, 2001b; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Whalen, 1991). In addition, allophonic variation in /t/ can augment or decrease lexical activation (Connine, 2004; McLennan, Luce, & Charles-Luce, 2003).

* Corresponding author. Fax: +1 319 335 0191.

E-mail address: bob-mcmurray@uiowa.edu (B. McMurray).

Some of the most striking evidence for sensitivity to sub-phonetic variation comes from studies examining within-category differences in lexically contrastive dimensions which were once believed to be perceived categorically, such as Voice Onset Time (VOT). VOT is defined as the time delay between the release of airflow blocked by the lips and the onset of glottal pulses from the closed vocal folds, and is the strongest cue for distinguishing between voiced plosives like /b/, /d/, and /g/ from voiceless sounds like /p/, /t/ and /k/. VOT is shorter for voiced than it is for voiceless sounds. A significant amount of variation is created by the surrounding phonetic context. Factors such as prosodic strength (Fougeron & Keating, 1997), speaking rate (Miller, Green, & Reeves, 1986), and place of articulation (Lisker & Abramson, 1964) all add variation to VOT. It varies consistently between speakers (Allen, Miller, & De Steno, 2003) and as a function of whether the segment arose from a speech error (Goldrick & Blumstein, 2006). This variability means that for a given segment and VOT there is some degree of uncertainty about the voicing category of that segment.

In a seminal study, Andruski, Blumstein, and Burton (1994) assessed cross-modal lexical priming as listeners heard fully voiceless targets or stimuli with 1/3 or 2/3 of the prototypical VOT value. Importantly, both the 1/3- and 2/3-voiced conditions represented within-category variants (reduced VOTs) that were nevertheless categorized reliably as tokens of the same voiceless phoneme. Andruski et al. found reduced priming in the 2/3 voiced condition (compared to the fully voiceless-condition), demonstrating that within-category variation in the signal can affect online word recognition. Utman, Blumstein, and Burton (2000) replicated these findings by showing greater priming for competitor words (e.g., *dime*, after hearing *time*). In several eye-tracking studies, McMurray and colleagues extended the Andruski et al. (1994) results by demonstrating that activation of target words and their competitors show gradient sensitivity to very small, within-category differences in VOT (McMurray, Aslin, Tanenhaus, Spivey, & Subik, in press-a; McMurray, Tanenhaus, & Aslin, 2002). Finally, a recent imaging study by Blumstein and colleagues reported activation to within-category differences in VOT in cortical areas believed to be sensitive to early speech processing (Blumstein, Myers, & Rissman, 2005).

The studies reviewed above clearly establish that word recognition shows sensitivity to fine-grained sub-phonetic detail. However, they do not establish whether within-category detail is short-lived or maintained for longer durations. If the spoken word recognition system maintains within-category detail, it could be useful for interpreting and integrating later arriving input. However, many models of speech processing, including models of phonetic and phonemic categories based on attractor dynamics (e.g., Damper & Harnad, 2000; Kuhl, 1991; McClelland & Elman, 1986; McMurray, Horst, Toscano, & Samuelson, in press-e; McMurray & Spivey, 1999), predict that initial gradient sensitivity to within-category detail is rapidly lost as the system gravitates toward an attractor state and a categorical representation of the phoneme. In simulations reported after Experiment 1, we show that, under most

circumstances, the influential TRACE model (McClelland & Elman, 1986) makes just this prediction.

The experiments reported here evaluate the time-course of sensitivity to fine-grained within-category differences in VOT by creating VOT continua for words such as *parakeet* and *barricade* ([bærəkɛɪd] and [pærəkɪt]). These words differ in initial voicing, and then overlap for multiple segments, before the point of phonemic disambiguation (the vowel in the final syllable). We used these stimuli to induce lexical garden-paths in which the final phonetic material is inconsistent with the preferred interpretation based on the initial segment (i.e., the non-words *barakeet* and *parricade*). If a continuous representation of VOT is available at the point of disambiguation, then the ease of recovery should be easier when the VOT is closer to the category boundary because the evidence for the preferred interpretation of the initial phoneme would be weaker than when the VOT is further from the category boundary. If gradient sensitivity is short-lived, such that the representation of the initial phoneme is more categorical, then the details of within-category VOT should have minimal effects on garden-path recovery.

The empirical literature does not establish whether sensitivity to sub-phonetic detail is retained across multiple phonetic segments or decays quickly, as predicted by attractor models. To the best of our knowledge, only a handful of studies have examined the time-course of sub-phonetic detail. Andruski et al. (1994) found that sub-phonetic differences as assessed in a priming task decayed within 250 ms. Using similar methods, Utman et al. (2000) found longer-lasting effects (at an ISI of 250 ms). However, both studies only found within-category sensitivity for a single token near the boundary, so it is unclear how long a gradient representation within the category could be maintained. McMurray, Tanenhaus, Aslin, and Spivey (2003) found evidence for sensitivity in visual fixations 1 s after word onset, but these effects could be due to fixations that were programmed in response to input from early in processing. All of these studies examined such effects only with single-syllable minimal-pair words in a decontextualized presentation. None of them examined whether this gradiency is preserved in a way that is functionally available to further word recognition processes.

Work by Luce and Cluff (1998) does suggest that the word recognition system might keep multiple options available for quite some time (which they characterize as a delayed commitment process). They demonstrated that offset-embedded words (e.g., *lock* in *hemlock*) are active well after the target word (*hemlock*) is unambiguously recognized, suggesting that the system can maintain alternative interpretations even after settling on a single word. However, this study did not examine representations below the lexical level—it is quite possible that sublexical representations (e.g., phonemes) settle on a fairly discrete representation (in accord with attractor-models), while lexical commitments operate over a longer time-course.

The most relevant studies were conducted by Connine and colleagues (Connine, 1987; Connine, Blasko, & Hall, 1991). Connine et al. (1991) created ambiguity at a phonetic level that could be resolved by later sentential context. They embedded a *dent/tent* VOT continuum in

sentences in which the subsequent context favored one interpretation over the other (e.g., “Because the *d*/tent on the pick-up was hard to find...” or “After the *d*/tent in the campground collapsed...”). Participants were instructed to identify the first phoneme of the target word as either /d/ or /t/. Results showed a small boundary shift in response to sentential context; contexts favoring a voiced interpretation yielded more /d/ responses than those favoring a voiceless context. Connine et al. (1991) found effects of context when the disambiguating information was 0, 1 or 3 syllables away from the target, and even when it crossed a clause boundary, but not with a 6- to 8-syllable delay. Connine et al. (1991) concluded that phonetic codes persist (and are available for integration with semantic material) for up to about a second, and more importantly that the system is capable of deferring a sub-lexical commitment (or at least modifying one) when the signal is phonetically and semantically ambiguous. Converging evidence comes from an unpublished study by Samuel (personal communication) who reports that phoneme restoration can be affected by subsequent sentential context.

While the Connine et al. (1991) study shows that ambiguous phonetic material can be retained for several syllables, the effect of the surrounding phonetic context was limited to VOTs near the category boundary. As we demonstrate in the simulations reported later, these are the only cases where TRACE successfully recovers from lexical garden-paths. Thus, it is possible that maintenance of phonetic detail is limited to phonetically ambiguous stimuli.

In sum, prior work makes a clear case for initial gradience, but is unclear about whether it persists (Andruski et al., 1994; McMurray et al., 2003; Utman et al., 2000). It demonstrates that lexical commitments are delayed (Luce & Cluff, 1998), but that they may be only malleable near the category boundary (Connine et al., 1991). All of this evidence is consistent with a model in which sub-phonetic detail is rapidly lost while lexical commitments are more graded.

Thus, the current studies examined whether sub-phonetic detail is rapidly lost, as predicted by TRACE, when the VOT is substantially further from the category boundary. We examined garden-path recovery for pairs of words such as *barricade*/*parakeet* and *bassinet*/*passenger* because they afforded us sufficient time after the initial consonant to measure the maintenance of within-category information. Each pair contained words that begin with an initial bilabial or coronal stop that differs in voicing, but otherwise overlap for at least four phonemes. We created a synthetic VOT continuum that spanned the target word and the cross-voicing non-word (*barricade* → *parricade*) in 5-ms steps. Word recognition was assessed using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). While eye movements are monitored, participants heard the word and selected its depicted referent from a computer display that contained five pictures: the two competitors (e.g., a *barricade* and a *parakeet*), two unrelated distracters, and a large red “X”. We added the X to the standard four-picture display introduced by Allopenna, Magnuson, and Tanenhaus (1998) to avoid forcing participants to choose among lexical alterna-

tives when they might have heard a non-word mispronunciation (e.g., *parricade* or *barakeet*).

Our study was designed to answer two questions. First, are listeners able to quickly revise their lexical hypotheses when subsequent information is inconsistent with a strongly preferred lexical bias? Our simulations with TRACE predict that revision will be difficult or impossible, except for tokens near the VOT category boundary. Second, and most importantly, we assess whether a gradient representation of the initial segment affects the revision process. If information about the VOT of the first phoneme is available when subsequent information that is inconsistent with the preferred lexical hypothesis arrives, then the time to recover from the “lexical garden-path” will be a function of VOT. As VOT approaches the category boundary (from the misperceived, “garden-path” end of the continuum), the time to revise the initial (provisional) lexical commitment should decrease.

The strongest test comes from those trials in which participants initially interpret the input as the competitor, and then revise this provisional interpretation after the disambiguating information arrives. We can then use look-contingent analyses to examine those trials in which participants are substantially down this garden-path. Thus trials can be identified as trials on which participants are fixating on the cross-category competitor (e.g., *barricade* when the auditory stimulus was *b/parakeet*) when the disambiguating information arrives (e.g., ...*keet*), on the assumption that what the participant is looking at is likely to represent his or her lexical hypothesis at that moment in time. For those trials, we can then examine the time it takes to arrive at the target, when (a) the next look is to the referent consistent with the last syllable (e.g., the *parakeet*) and (b) the participant selects that picture as the referent. If the system preserves gradient information about VOT, the time to launch a corrective saccade should be a function of the distance of the initial VOT from the category boundary.

Experiment 1

Methods

Participants

Eighteen University of Rochester undergraduates served as participants in this experiment. All reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA ethical guidelines. Testing was conducted in two sessions lasting approximately 1 h and participants were compensated \$10 for each session. Data from one participant who did not return for the second session were excluded.

Auditory stimuli

Experimental materials consisted of 10 pairs of phonemically similar words, each beginning with labial or coronal consonants (6 b/p pairs and 4 t/d pairs). Pairs were selected that differed in initial voicing of the first consonant, but were identical for the next 3–5 phonemes. Table 1 lists these pairs,

Table 1
Experimental items used in experiments 1–3

Voiced	Voiceless	Point of disambiguation (phonemes)	Point of disambiguation (ms)
Bumpercar	Pumpnickel	6	240
Barricade	Parakeet	6	270
Bassinet	Passenger	6	335
Blanket	Plankton	6	225
Beachball	Peachpit	5	280
Billboard	Pillbox	5	210
Drain Pipes	Train Tracks	6	285
Dreadlocks	Treadmill	5	170
Delaware	Telephone	5	225
Delicatessen	Television	5	165

Point of disambiguation refers to the index of the phoneme at which the two words could be differentiated (assuming ambiguous initial voicing).

including the amount of overlap in both number of phonemes and in milliseconds. Two 10-step (0–45 ms) VOT continua were created from these pairs (one for each word).¹

VOT continua were synthesized with the KlattWorks (McMurray, in preparation) interface to the Klatt (1980) synthesizer, modeled on tokens of natural speech. Several recordings of each pair were made by an adult male speaker in a quiet room. From these recordings, tokens were selected that best matched each other for speaking rate, voice quality and for which automated formant analysis yielded the best results. After selecting these pairs, pitch and formant tracks were extracted automatically using Praat (Boersma & Weenink, 2005) and used as the basis for the *F0–F5* parameters of the synthesizer.

For each pair, synthesis was based on the voiced endpoint (e.g., *barricade*). Formant frequencies were based on the measurements made with Praat. The remaining parameters (voicing [AV], aspiration [AH], frication [AF] and formant bandwidths [B1–B6]) were selected to best match the spectrogram and envelope of the original recorded token. Several recorded tokens had non-zero VOTs (5–10 ms). For these, the VOT was set to 0 ms by adjusting the onset of voicing to be temporally coincident with the release burst.

After constructing a natural-sounding voiced token (e.g., *barricade*), the voiceless competitor (e.g., *parakeet*) was created in three steps, adopting methods chosen to ensure maximally natural-sounding synthetic speech that was completely identical for the overlapping segments within each pair. First the parameters of the voiced token were copied. Second, the VOT of the token was set to 45 ms by setting the onset of voicing to 0 and aspiration to 60 dB for the first 10 frames (45 ms). Next, parameters that occurred after the point of disambiguation (e.g., /e/ in the final syllable of *barricade*) were modified so that the remainder of the word best matched the voiceless token by comparing the spectrograms of the synthetic and

natural versions. The stimuli are available from the first author.

Having synthesized the two tokens for each of the 10 pairs, twenty 10-step VOT continua were constructed (one from each token). To construct continua from the voiced words (e.g., *barricade*), the onset of voicing was cut back in 5-ms increments and replaced with 60 dB of aspiration. To construct continua from the voiceless words (e.g., *parakeet*), the onset of voicing was decreased in 5-ms increments, and the duration of aspiration was decreased accordingly. These procedures guaranteed that for a given VOT, targets and competitors were parametrically and acoustically identical until after the point of disambiguation (POD: /eet/ vs. /ade/). After synthesis, the POD was measured directly from the synthetic stimuli. The mean POD for the set was 240 ms or 4.5 phonemes after word onset (Table 1). KlattWorks scripts for each of these continua are available on the web at [JML online archive].

In addition to these 20 VOT continua, 10 pairs of filler items were also synthesized (Table 2). Filler items began with continuants and fricatives so as to be minimally overlapping with the target continua, and the filler set was roughly analogous to the distribution of voiced and voiceless labials and coronals in the target set (four /l/- and /s/-initial items and six /r/- and /f/-initial items). Filler items were not phonetically similar to each other and had minimal overlap. Filler items were synthesized similarly to target items using automatically extracted pitch and formant frequencies, and by matching synthetic spectrograms to the spectrogram of the recorded items.

We assumed that participants would perceive some of the test stimuli as mispronounced, especially for stimuli taken from opposite ends of the VOT continuum (e.g., *barricade* with a VOT of 45 might be heard as the non-word *parricade*). In order to minimize the likelihood that mispronunciation would be a cue that distinguished the target/competitor pairs from the filler pairs, we created a mispronounced version of each filler item: /l/-initial items were “mispronounced” as /r/-initial (e.g., *rimousine*); /r/-initial items were mispronounced as /l/-initial (*lrestaurant*); /s/-initial items were mispronounced as /f/-initial (*faxophone*); and /f/ items were mispronounced as /s/-initial (*sotograph*). Note that /l/ items never appeared with /r/ items (/l/ was paired with /s/), and /f/ items did not appear with /s/ items, so the mispronunciation did not introduce competition among the displayed alternatives due to phonetic similarity. During the course of the experiment, the correctly pronounced version of the filler items was heard on 75% of the filler trials, and the mispronounced version on the other 25%.

Table 2
Filler items used in Experiment 2

/l/	/r/	/s/	/f/
Lemonade	Restaurant	Saxophone	Photograph
Limousine	Rabbit	Secretary	Fountain
Lobster	Reptile	Sunbeam	Factory
Lantern	Raspberry	Spiderweb	Farmyard
	Referee		Fireplace
	Rectangle		Footstep

¹ Note that *barricade* and *parakeet* were both synthesized with primary stress on the initial syllable (in some dialects of English, *parakeet* has secondary stress on the initial syllable). While the first syllable of *bassinet* receives secondary stress (and *passenger*'s has primary stress), the first syllable of both continua were acoustically identical.

Visual stimuli

Visual stimuli were 40 color drawings corresponding to the 20 target/competitor pairs and the 20 filler items. For each item, several pictures were downloaded from a large commercial clipart database. One picture was selected by groups of 3–4 viewers as being the most representative, easiest to identify, and least similar to the others in the complete set. In a few cases, images were edited to remove extraneous components or to alter colors.

Procedure

When participants arrived in the laboratory, informed consent was obtained and the instructions were given. An Eyelink II eye tracker was calibrated using the standard 9-point calibration procedure. The experiment was programmed using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, participants saw five pictures in a pentagonal formation on a 20" computer monitor. Each picture was equidistant from the center of the screen and was 250 × 250 pixels in size. While the five screen locations did not change, the ordering of the pictures within these locations was randomly selected for each trial. These pictures corresponded to the target (e.g., *barricade*), the competitor (e.g., *parakeet*), and the two matching filler items (e.g., *lemonade* and *restaurant*) that had been randomly selected (for that participant) to form a set of four. The fifth picture was a large red X that participants were instructed to select when they heard a mispronounced word.

Each trial began with a display of the pictures and a small blue circle, which was presented in the center of the screen. After 750 ms this circle turned red, signaling the participant to click on it with the computer mouse. The red circle then disappeared and one of the auditory tokens was played. Participants clicked on a picture, which ended the trial. There was no time-limit on the trials, but subjects typically responded in less than 2 s ($M = 1601$ ms, $SD = 166.7$). Eye movements were only analyzed before and after the POD of the stimulus ($M = 240$ ms), so very late mouse-click responding was unlikely to affect the eye-movement data.

Design

Experiment 1 made use of 10 target/competitor pairs, each of which was composed of 10-step VOT continua. In addition, pairs were designed such that either word could serve as the spoken target. Thus on some trials the voiced word (*barricade*) was the target and the competitor was voiceless. On other trials the voiceless word (*parakeet*) was the target and the voiced word (*barricade*) was the competitor. With an equal number of filler trials, this yielded 10 (items) × 2 (voiced vs. voiceless target) × 10 (VOTs) × 2 (fillers) = 400 trials for a single repetition of the design. We estimated that seven repetitions of each VOT would be needed for adequate statistical power, yielding a total of 2800 trials—far too many for a participant to complete. Thus, we adopted a Latin-square design in which each participant was randomly assigned five of the 10 item-sets. There were no constraints on this randomization—each subject's 5 item-sets were chosen

independently of the others, although the distribution was roughly equivalent across subjects.² This led to 700 experimental trials (5 items × 2 target-types × 10 VOTs × 7 reps). We further reduced the number of fillers to slightly less than the number of experimental trials (632). This yielded 1332 total trials per participant, which were administered in two sessions, each lasting approximately 1 h.

To maintain consistency with prior work (McMurray et al., in press-a), filler items were consistently paired with experimental items by initial consonant. If fillers had been selected at random on each trial, participants might have noticed a relationship between the voicing pairs (since, for example *barricade* is consistently present with *parakeet*, but not with the other two items). Thus, sets of four items were randomly selected from each category, and were paired consistently throughout the experiment. Each set (randomly determined for each participant) consisted of one continuant (/l/ or /r/) and one fricative (/s/ or /f/). B/p pairs were paired with /r/- and /f/- initial fillers. D/t pairs were paired with /l/- and /s/-initial fillers. Pairs were randomly assigned with the exception that filler items could not be semantically related to each other or to the target/competitor pairs.

Eye-movement recording and analysis

Eye movements were recorded at a 250-Hz sampling rate using an SR Research Eyelink II eyetracker. This tracker uses the pupil and corneal reflection to determine eye-position, and a set of infrared lights mounted on the corners of the computer monitor to compensate for head-movements. Moment-by-moment eye-position coordinates were automatically parsed into saccades, fixations and blinks, using the conservative default settings. Saccades and the subsequent fixation were then combined into "look". A look began at the onset of a saccade and ended at the offset of the ensuing fixation. All analyses reported here were conducted on the basis of these looks.

The Eyelink II system shows an inherent drift in its report of eye-gaze. To counter this, the drift correction procedure was run every 40 trials. In addition, the area-of-interest corresponding to each picture was expanded by 40 pixels to relax the criterion for what counted as a fixation. This did not result in any overlap between the areas-of-interest of the five pictures. Thus any look for which the fixation fell within this expanded area-of-interest was counted as a look to the corresponding picture.

² Barricade and Delicatessen were slightly oversampled (with 12 subjects being assigned to these continua), while blanket and bassinet were undersampled (6 and 5 subjects, respectively). Other continua were between 7 and 10. This creates slightly unequal sample sizes between cells in the item analyses, however, we point out that within an item all levels of the factor in question (VOT) had the same sample size. Moreover, item analyses here play a somewhat secondary role in that the primary factor (VOT) covaried with item (e.g., each item was head at each VOT), so item is not confounded with our experimental manipulation.

Results

Three separate sets of analyses were performed. The first set of analyses was conducted to confirm that our stimuli produced initial gradient effects of VOT, replicating the pattern found in McMurray et al. (2002, in press-a) because finding this pattern was a prerequisite for addressing questions about time-course. The second examined the pattern of phoneme identification (mouse-click) responses to assess whether participants' overt labeling demonstrated an ability to recover from initial ambiguity. In particular, it asked if participants clicked the target consistent with the end of the stimulus, or did they generally garden-path and classify the stimuli as non-words (or the competitor). The third analysis addressed our primary question: do differences in VOT systematically affect ambiguity resolution? This analysis examined trials in which participants' initial fixation was directed to the competitor and measured the latency to fixate the target.

Eye-movement evidence for initial gradiency

On each trial, we limited our analysis to the last fixation prior to the point of disambiguation (plus 200 ms to account for saccadic planning). From this segment of each trial, the proportion of fixations to the voiced and voiceless target was computed for each participant, at each VOT. This was then converted to a measure of bias by subtracting the proportion of fixations to the voiceless target from the proportion of fixations to the voiced target. Thus, a value of 0 would indicate that participants were equally likely to fixate both, positive values would indicate they were biased towards a voiced interpretation and negative values would indicate bias towards a voiceless interpretation.

Fig. 1 shows that as the VOT increased from 0 to 45 ms participants were increasingly more likely to commit to the voiceless interpretation, replicating the gradient effects of VOT reported by McMurray et al. (2002, in press-a). This did not appear to vary as a function of the lexical-endpoint of the continuum—the effect was the same whether it was

the *barricade/parricade* or the *parakeet/barakeet* continuum. This pattern of results is consistent with the fact that these initial eye movements were generated prior to the POD and well before the offset of the word.

Statistical support for these patterns came from an ANOVA examining this bias measure as a function of VOT and the lexical-endpoint (whether the lexical-endpoint of the continuum was voiced [*barricade*] or voiceless [*parakeet*]). Here, we expected an effect of VOT, but no effect of lexical-endpoint—since these eye movements were launched prior to the point of disambiguation, listeners should not know what the word was (Table 3 for statistical results referenced by row). There was a significant main effect of VOT (row 1), with more fixations directed to the voiceless target as VOT increased. This relationship took the form of a linear trend (row 2). Lexical-endpoint was not significant (row 3) and did not interact with VOT (row 4).

While the trend analysis demonstrates that the linear (gradient) model provides a good description of the relationship between VOT and initial commitment, we cannot conclude from the trend alone that it is a better fit than other non-linear functions. Of greatest concern, of course, is the possibility of a categorical step function at the

Table 3

Statistical tests examining effect of VOT and target-type on the proportions of fixations to the competitor in Experiment 1

	Factor		<i>df</i>	<i>F</i>	<i>P</i>
1	VOT	<i>F</i> ₁	9, 144	16.9	<.0001
		<i>F</i> ₂	9, 81	6.5	<.0001
		min <i>F</i>	9, 144	4.7	.0001
2	VOT (trend)	<i>F</i> ₁	1, 16	46.3	<.0001
		<i>F</i> ₂	1, 9	10.2	.01
		min <i>F</i>	1, 13	8.4	.013
3	Lexical-endpoint	<i>F</i> ₁		<1	
		<i>F</i> ₂		<1	
4	Endpoint × VOT	<i>F</i> ₁	9, 144	1.6	>.1
		<i>F</i> ₂	9, 81	1.8	.13
		min <i>F</i>	9, 215	<1	

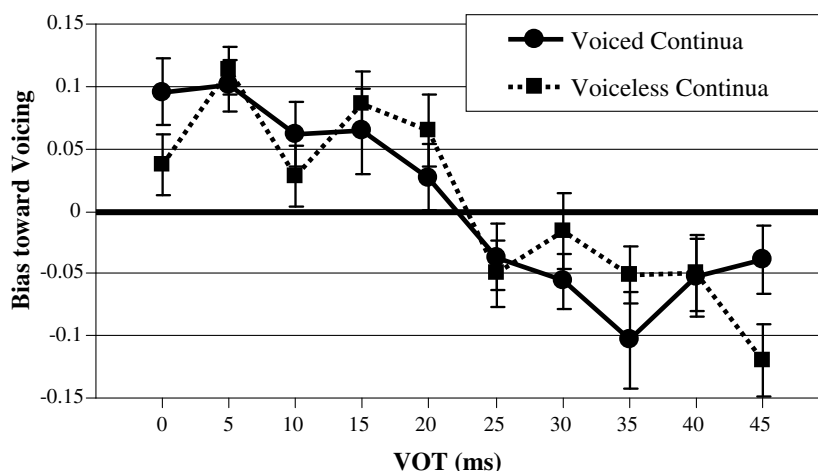


Fig. 1. Degree of bias toward voiced or voiceless competitors prior to the point of disambiguation as a function of VOT in Experiment 1.

boundary. Additionally, even if a linear model is a better fit overall, it may still derive from an underlying categorical relationship if there is variability between participants in the location of the category boundary. Thus, we conducted an additional analysis to more firmly establish the gradient of the initial commitment.

In this analysis, two models were compared to determine the best fit to the data (see Appendix B for details). In the first (linear) model, the relationship between increasing VOT and bias toward the voiceless interpretation was modeled as a linear function. In the second (categorical) model, this relationship was modeled as a step function approximated by a three-parameter logistic function. In this function the lower asymptote, the upper asymptote and the crossover point (category boundary) were free to vary, but the slope was fixed at a very steep value to approximate a step function. Both models were simple mixed effects models in which each participant's data was fit to the linear and logistic functions separately (to avoid issues of between-subject category boundary variability). Thus, the linear model as a whole had separate slopes and intercepts for each subject, while the logistic model had separate upper asymptotes, lower asymptotes and crossover points for each subject.

The non-linear model should in principle be superior to the linear model because it has one extra degree of freedom (for each subject). This makes it difficult to directly compare models. However, the Bayesian Information Criteria (BIC) measure (Schwarz, 1978) was designed to facilitate this comparison (when both models are fit to the same data), by taking into account the number of parameters and the sample size when evaluating a goodness-of-fit measure (in this case log-likelihood). The model with the lower BIC is the better fit.

Both models were fit to a dataset consisting of the voicing-bias measure averaged at each VOT and for each continuum for each subject. Both fits were good, with the linear model accounting for 11.1% of the variance and the logistic model accounting for 11.8%. However, the BIC measure strongly favored the linear model, yielding a BIC of 68.8 for the linear model and 180.2 for the categorical model. Moreover, an analysis of the BIC for individual subjects revealed that 15 of the 17 subjects showed lower BIC values for the linear than the logistic model. Thus, we can safely rule out a non-linear (categorical) model as the underlying form of the function relating time-to-target-disambiguation to VOT.

Mouse-click evidence for recovery

Participants chose the target that was consistent with the final disambiguating segment 83% of the time for the voiced targets and 70.5% of the time for the voiceless targets. These sub-asymptotic percentages can be attributed to the ambiguous initial segment and were largely due to increased non-word responding when the VOT indicated an initial phoneme that conflicted with the target (e.g., *barricade* with a VOT of 45 ms). When non-word (X) responses were excluded, these figures increased to 99.4% and 98.6%, respectively—subjects rarely chose the competitor. Moreover, when the VOT of the initial consonant was consistent with the correct pronunciation (e.g., *barricade* with a VOT

of 0), participants responded correctly to 92% and 86% of the items from the voiced and voiceless regions of the continua, respectively.

Fig. 2 shows the percentage of trials in which participants clicked the voiced target (B/D), the voiceless competitor (P/T), the X, or the fillers for trials in which the target was voiced (panel A) or voiceless (panel B) as a function of VOT. These data show that identification is systematically related to VOT: as the VOT departs from the endpoint (0 ms for voiced words, 45 ms for voiceless words), identification responses to the target decrease and identification responses to the non-word increase.

These results also indicate that participants' final judgment about the identity of the target word was not irretrievably influenced by the initial consonant. Participants rarely clicked the competitor word (e.g., *parakeet* when the target was from the *barricade* continuum), averaging 0.4% ($SD = 0.4%$) for the voiced continua (e.g., those based on *barricade*) and 0.6% ($SD = 1.2%$) for the voiceless (e.g., those based on *parakeet*). Thus, true, irrecoverable garden-paths were rare.

The more likely response given misleading onsets was to click the X (non-word). However, as shown in Fig. 2A, even when the onset consonant was maximally discrepant from the target word (e.g., it was incorrectly pronounced as *parricade* with a VOT of 45 ms), participants recovered from the phonetic mismatch and selected the target picture (the *barricade*) that was consistent with the disambiguating information. Moreover, 10 of the 17 participants almost never used the X response category: their average rate of target responding across the entire VOT continuum was 95.6% ($SD = 4.5%$), and they averaged 88.9% target responding on the maximally distal VOT ($SD = 13.4%$). Thus, these 10 participants were consistently able to recover from a mismatch in the initial consonant at VOT values well beyond the category boundary.

A similar pattern was observed for the voiceless continua (Fig. 2B), with participants selecting the picture of the *parakeet* far more than the *barricade* or the X overall. Even for the most extreme mismatch (e.g., *barakeet* with a VOT of 0 ms), participants selected the *parakeet* 51.6% of the time ($SD = 38.5%$). Moreover, 9 of the 17 participants almost never used the X response category: their average rate of target responding across the entire VOT continuum was 90.9% ($SD = 12.4%$), and was 84.8% at the most extreme VOT ($SD = 16.0%$).

Finally, the overall pattern of identification results was not consistent with a sharp categorical boundary of any kind—identification functions transitioned slowly and smoothly towards non-words. Thus, the identification results demonstrate that many participants were able to fully recover from a lexical garden-path and all participants were able to recover partially. Moreover, recovery appears to be a gradient function of VOT. A stronger test of the gradiency hypothesis involves an assessment of how much temporary activation is elicited by the garden-path information from the “incorrect” initial consonant. Such a test cannot be provided by mouse-click judgments alone, but rather requires an online assessment such as eye movements to the pictured alternatives.

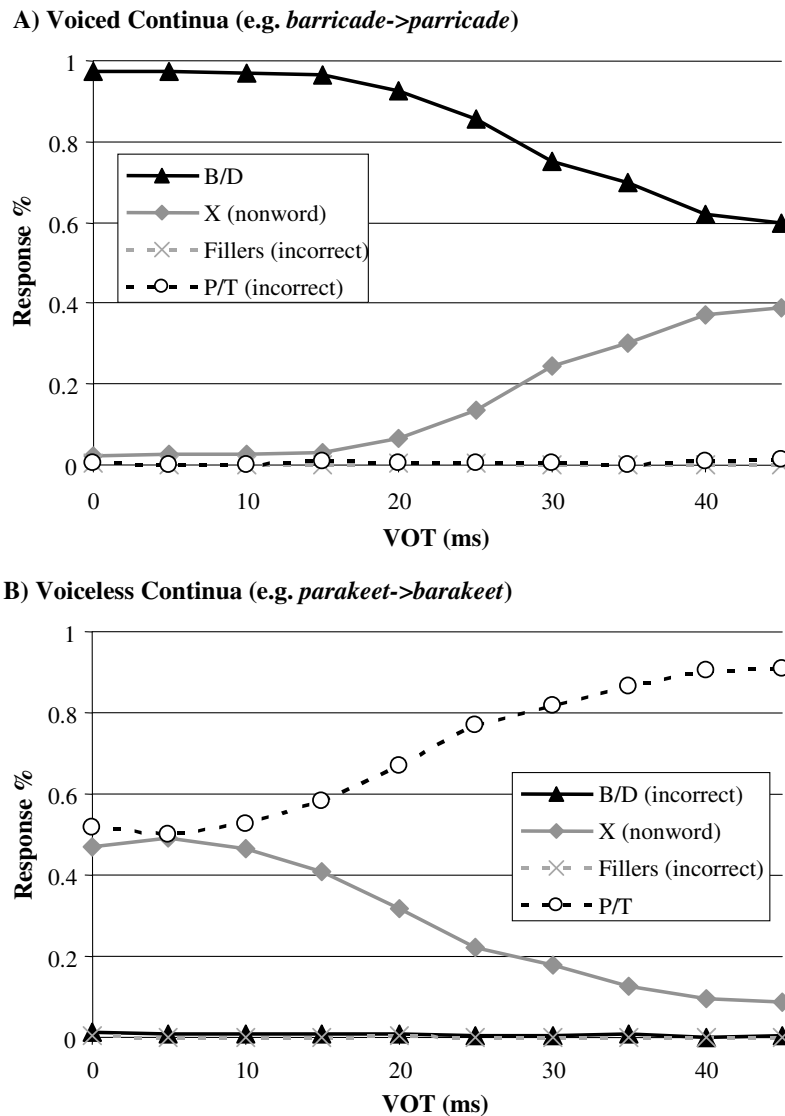


Fig. 2. Identification results for Experiment 1. (A) Identification for voiced targets (e.g., *barricade*). (B) Identification for voiceless targets (e.g., *parakeet*).

Eye-movement evidence for gradient recovery

In order to assess our hypothesis that VOT would systematically affect listener's ability to resolve the initial ambiguity, two series of analyses based on the sequence of fixations were conducted. Trials were included in the first series of analyses if (1) the participant was fixating the competitor prior to the POD (plus the 200 ms of oculomotor planning time) and (2) the next fixation after the POD was directed to the target. For these trials, the time between the POD and the fixation to the target was the dependent variable [time-to-target]. Unfortunately, these criteria led to a fairly sparse dataset. On average participants had 2.55 trials contributing to each VOT, but there was high variance in the number of trials (VOT: $SD_{\text{within-participant}} = 1.31$; $SD_{\text{between-participant}} = .77$), yielding 60 empty cells out of the 340 cell ANOVA (2 target-types \times 10 VOTs \times 17 participants).

To deal with the sparse data problem, a hierarchical regression analysis was conducted by treating VOT as a continuous covariate (Fig. 3A). In this analysis, VOT was recoded as distance from the word-endpoint of the continuum, which we term relative VOT (rVOT). Thus, a 0-ms VOT in the *b/parricade* continuum was coded as an rVOT of 0, while a VOT of 0 in the *b/parakeet* continuum was coded as an rVOT of 45. This meant that for both voiced (*b/parricade*) and voiceless (*b/parakeet*) continua, we predicted an increase in time-to-target as rVOT increased. In addition, whether the continuum had a voiced or voiceless target (lexical-endpoint), as well as the interaction, were included as covariates. We did not expect that the lexical-endpoint would influence recovery (as this would suggest, for example, that participants were faster overall to recover in voiced continua or voiceless continua), nor did we expect an interaction. This type of analysis more grace-

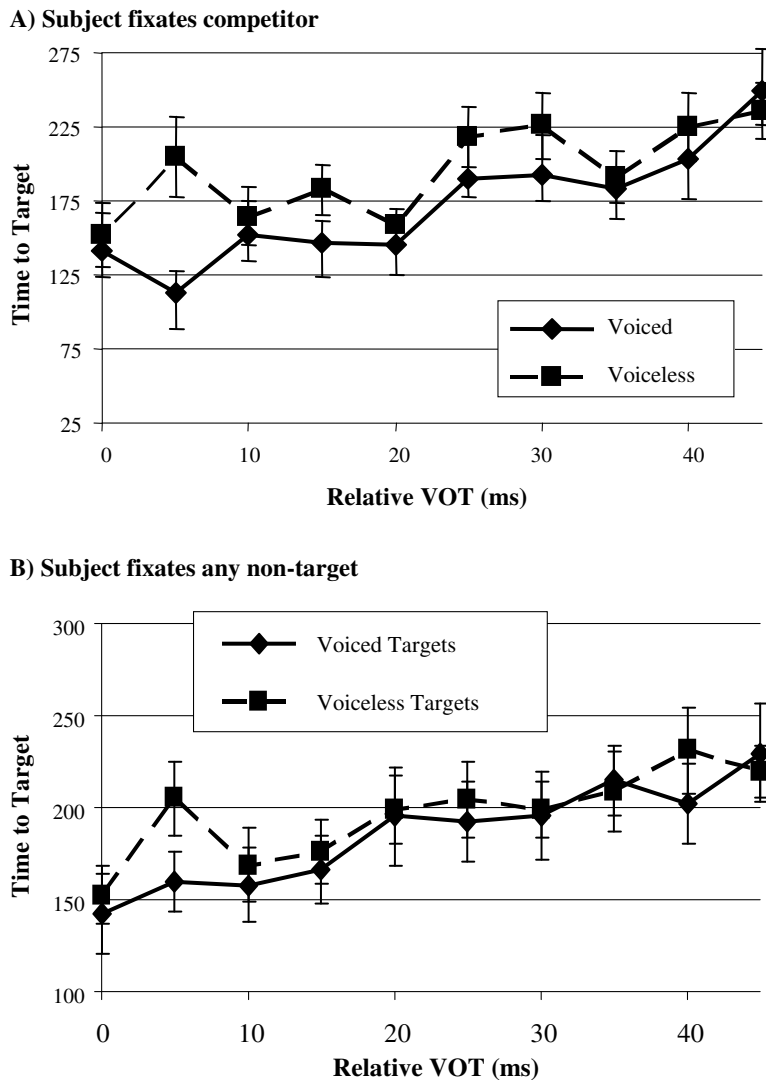


Fig. 3. Time to fixate the target after the point of disambiguation as a function of distance from the lexical-endpoint (rVOT) for continua based on voiced and voiceless targets. (A) time-to-target for trials in which the participant was fixating the competitor just before the point of disambiguation. (B) Time-to-target for trials in which the participant was fixating any non-target.

fully copes with missing data because it treats VOT as a single continuous covariate rather than a set of 10 independent cells. Complete results for these analyses are presented in Table 4.

On the first step of the regression, participant codes were added to the model and found to account for 13.9% of the variance. In the second step, rVOT and lexical-endpoint accounted for an additional 13.1% of the variance. Lexical-endpoint was significant individually, with participants fixating the target approximately 24 ms faster for voiced continua than for voiceless continua ($B = 24.4$, 95% $CI_{\text{slope}} = 17.6$ ms). This asymmetry is likely due to the fact that our continua appeared to have boundaries that were not centered between 0 and 45 ms of VOT—more voiced sounds were perceived than voiceless. Thus, there was a token-frequency bias to recover to voiced sounds, which could result in the significant effect of endpoint. Crucially,

time-to-target was significantly related to rVOT ($B = 1.98$ ms/VOT, 95% $CI_{\text{slope}} = .61$). As the VOT departed from the lexical-endpoint, participants took longer to switch from the competitor to the target. Finally, the interaction term was added to the regression and accounted for no additional variance.³ Thus, the effect of rVOT was not different for the two types of continua.

A second regression analysis that used items as the fixed effect revealed similar effects (also in Table 4). On

³ It is possible that these seemingly graded results arose from a discrete system in which time-to-target was stochastically selected from one of two discrete categories (target or garden-path), and the probability of choosing one of the distributions varied with VOT. To rule this out, we conducted an additional analysis of the frequency distribution of the time-to-target values. No evidence for a bimodal distribution was seen. The effect of VOT was best described as a linear change in the mean of a unimodal distribution.

Table 4

Results of a regression analysis examining the effect of VOT and lexical-endpoint on time-to target in Experiment 1

Analysis	Step	Factor	R^2_{change}	df	F/t	P
Subject analysis	1	Subject	.139	15, 263	2.6	.001
	2	Main effects	.131	2, 261	23.5	<.001
		Lexical-endpoint		261	2.7	.007
		rVOT		261	6.3	<.001
3	Interaction	.006	1, 260	2.3	>.1	
Item analysis	1	Items	.054	9, 171	1.1	>.1
	2	Main effects	.237	2, 169	28.3	<.0001
		Lexical-endpoint		169	3.3	.001
		rVOT		169	6.7	.001
	3	Interaction	.001		<1	

the first step, item codes accounted for 5.4% of the variance but did not reach significance, suggesting that variance in time-to-target may be due more to between-participant factors (e.g., generalized reaction time) than item-specific factors. When rVOT and lexical-endpoint were added in the second step, an additional 23.7% of the variance was accounted for. As before, lexical-endpoint was significant in that participants were faster to fixate the target for voiced continua ($B = 31.3$ ms, 95% $CI_{\text{slope}} = 18.4$ ms). Importantly, rVOT was also significant ($B = 2.13$, 95% $CI_{\text{slope}} = .632$). Finally, the interaction term did not account for any additional variance on the third step ($R^2_{\text{change}} = .003$).

Follow-up analyses were conducted in which the inclusion criteria were relaxed to allow any trial in which the participant was fixating any non-target object prior to the POD to be included. These less restrictive criteria allowed more data to contribute to the analyses ($M = 7.4$ trials per cell), and were necessary for comparison with Experiment 2. Because there are now no empty cells, this enabled the use of a 10 (rVOT) \times 2 (voiced or voiceless target) ANOVA to assess effects of VOT on time-to-target (Table 5). Mean time-to-targets at each VOT for this analysis are shown in Fig. 3b. The effect of lexical-endpoint was not significant under this analysis, suggesting that the effect seen in the first analysis may have been relatively small (row 1). However, rVOT was highly significant (row 2). This was the result of a significant linear trend (row 3): as VOT departed from the value of the lexical-endpoint, time-to-target systematically increased. rVOT did not

interact with lexical-endpoint (row 4), suggesting that this effect was not different for voiced and voiceless targets.

Gradient or categorical recovery?

As in our analysis of the initial commitment, we again needed to determine that the linear relationship seen in the prior analysis was a better fit than a non-linear (categorical) function, once variability in category boundary was accounted for. Thus, we used a similar procedure as before, fitting two mixed effects models to the data. In these models the linear or logistic functions described the relationship between VOT and time-to-target. Unlike the prior analysis, however, the time-to-target dataset was already limited to trials during which the participant was fixating particular objects prior to the point of disambiguation. Thus, dividing the data further by participant left us with relatively small dataset for each fit. As a result, we used the data from the second analysis, which computed time-to-target for trials in which the participant was fixating any non-target object prior to the point of disambiguation. This yielded, on average, 148.3 data-points per participant ($SD = 58.5$).

Results strongly supported a gradient model over a categorical model. In absolute terms, both models fit the data quite well, although the linear fit was better (Linear: $R^2 = .199$; Logistic: $R^2 = .120$). However, the BIC measure showed a strong advantage for the linear model over the logistic model (Linear: 32,133; Logistic: 32,375), and every participant had a lower BIC score for the linear model than the logistic model. Thus, we can conclusively rule out a categorical model as the underlying form of the function relating time-to-target to VOT.

Table 5

Results of an ANOVA examining time-to-target as a function of lexical-endpoint and rVOT in Experiment 1

	Factor		df	F	P
1	rVOT	F_1	9, 144	6.5	<.001
		F_2	9, 81	3.9	<.0001
		min F	9, 172	2.4	.012
2	rVOT (trend)	F_1	1, 16	35.6	<.001
		F_2	1, 9	16.0	.003
		min F	1, 17	11.0	.004
3	Lexical-endpoint	F_1	1, 16	2.4	>.1
		F_2		<1	
4	Endpoint \times rVOT	F_1		<1	
		F_2		<1	

Discussion

During the temporal interval prior to phonemic disambiguation, fixations to the voiced and voiceless target showed gradient sensitivity to VOT, replicating earlier results and demonstrating that participants make provisional rather than categorical commitments. Importantly, participants were able to revise their initial interpretations when they encountered new phonetic evidence after an ambiguous initial segment. All participants showed an ability to identify some of the garden-path stimuli (e.g., *parricade* as a poor exemplar of *barricade*), although there were individual differences in how much of a mismatch

was tolerated. Most crucially, the continuous-valued VOT of the initial segment was linearly related to time to recover from the initial interpretation, suggesting that fine-grained information about VOT was retained over the 240-ms period prior to the POD.

We have argued that models that incorporate attractor dynamics at the phoneme level are likely to predict short-lived sensitivity to within-category differences in VOT. To illustrate this claim we conducted simulations using TRACE, as an example of an attractor-based model, because it successfully simulates a wide range of effects in spoken word recognition (for recent review see Gaskell, 2007), and because it includes quasi-featural level information that can be used to simulate sensitivity to sub-phonetic detail.

TRACE predicts that as a word unfolds over time, multiple lexical candidates become activated, including words that overlap at onset and words that rhyme (Alloppenna et al., 1998). TRACE also exhibits initial fine-grained sensitivity to sub-phonetic detail, with activation for lexical competitors being proportional to continuous variation at the feature-level. For example, TRACE successfully simulates the pattern of results presented in McMurray et al. (2002, in press-a) for monosyllabic words such as *p/beach* (we are using *p/b* as short hand for a VOT continuum). The question here, then, is whether TRACE can preserve this initial sensitivity to enable recovery from subsequent mismatching information.

TRACE simulations

All simulations were conducted using the jTRACE simulator (Strauss, Harris, & Magnuson, 2007). Stimuli were analogous to Experiment 1 (though they were limited by TRACE's phonetic inventory), consisting of three pairs of temporarily ambiguous words: *barricade/parakeet*, *billboard/pillbox* and *beachball/peachpit*. Each word-pair was used to create 9-step VOT continua. Each pair was yoked with two fillers (an *l*- and *s*-initial multisyllabic word). Each simulation was run for 200 epochs, with the feature-spread parameter set to the default (6). This means that there were six frames between the peak activation for each phoneme's featural input. Thus, the POD for the *barricade/parakeet* and *beachball/peachpit* continua was at frame 30, and for *billboard/pillbox* at frame 24.

We first simulated the results from Experiment 1 using the standard TRACE parameters. Fig. 4 reports TRACE simulations for *parakeet* and *barricade* showing gradient effects of VOT near the category boundary, but an inability to recover from the initial mismatch when the stimulus was one-step over the category boundary. In fact, in panel C, the competitor (e.g., *parakeet* after hearing *parricade*) is nearly as active as *barricade* in panel A. Fig. 5 shows the final activation values for each word in those same simulations as a function of VOT. In a sense, this reveals the model's final decision (analogous to the mouse-response data). At the disambiguating phoneme, TRACE only shifts its preferred interpretation when the initial VOT is at the category boundary, thereby showing little capacity to recover from garden-path effects. This suggests a contrast

with the mouse clicking results of Experiment 1. We did not observe that nothing was active in the model (e.g., a "non-word" response). Rather, the model has fully garden-pathed to the competitor, something we did not observe in subjects' responding.

In order to determine the conditions under which TRACE could or could not simulate the data from Experiment 1, we then manipulated a number of parameters that control the dynamics of activation flow in TRACE. In particular we sought to determine if there are any instantiations of TRACE which can defer a commitment long enough to recover from the garden-path. For ease of comparison with the empirical data, TRACE activations were converted to fixation proportions using the Luce-Choice rule, as in previous simulations of eye-movement data from TRACE (Alloppenna et al., 1998; Dahan, Magnuson, & Tanenhaus 2001a; Dahan et al., 2001b; see Appendix A for details), although results were similar when the raw activations were examined.

The use of the Luce-Choice rule as a linking hypothesis provides a transformation of model activation that can be compared to empirical results from studies examining the proportion of trials on which the participant is fixating on each target at any point in time. The empirical data for Experiment 1 are shown in Fig. 6. As shown in the figure, participants correctly identified the target (all of the curves asymptote at .8—participants looked at the target 80% of the time regardless of VOT). It also shows that this process was systematically affected by the initial VOT, in that the rise-time to the target is delayed in the non-prototypical VOTs, and that this effect is small and graded with distance from the category boundary.

Given this ability to make a more detailed comparison between model and data, we next sought to determine which versions of TRACE (if any) could successfully recover from the mismatch, and show gradient sensitivity to the initial phoneme. There are 19 parameters that control the activation flow of TRACE, making an exhaustive search of the parameter space impractical. Therefore we examined a variety of parameters that would control the speed at which the model commits to a single word, as well as parameters that affect its ability to retain alternatives over time. These included (a) feedback from word nodes to phoneme nodes; (b) lateral inhibition at the phoneme level and the word level, (c) rate of decay of activation and (d) growth of activation. The complete set of parameters that were manipulated (along with their default and manipulated values) is provided in Table 6.

Fig. 7 shows the results of a number of representative parameter manipulations. Each panel displays the predicted fixation proportions as a function of time, for each continuum step, and for each parameter of TRACE that was manipulated. In order to simplify the analysis, the continua with voiced targets (e.g., *b/parricade*) were combined with the voiceless counterparts (*b/parakeet*) by recoding the continuum step as distance from the target (e.g., a fully voiceless token became +9 for the *b/parricade* continuum, and +0 for the *p/parakeet* continuum). Panel A shows a synopsis of the eye-tracking data from Fig. 6 for ease of comparison.

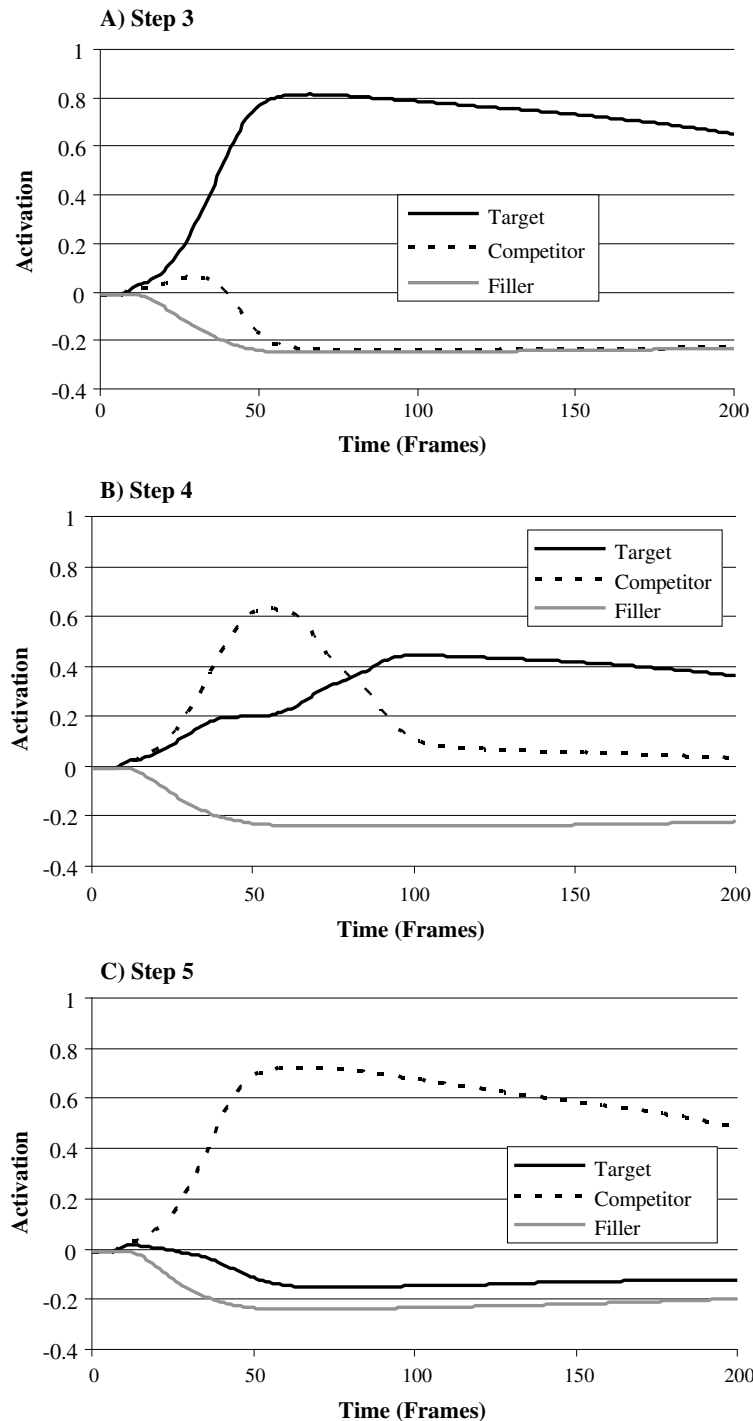


Fig. 4. Activation for the target (*barricade*), competitor (*parakeet*) and fillers as a function of time in TRACE given a 9-step VOT continuum. (A) Step 3, a voiced sound adjacent to the category boundary. (B) Step 4: near the category boundary. (C) Step 5, a voiceless sound just over the category boundary.

Fig. 7B displays the results for the default parameter set. The model can correctly identify the target at +0 through +3 steps from the lexical-endpoint, but it consistently fails to activate the target at steps +5 through +8. When word-to-phoneme feedback was eliminated (panel C), results

were similar—the only difference was that the model could reactivate the target at step +5 for *peachpit* and *billboard*, but not for the other continua (hence the asymptote at .4).

We next examined the parameters that control lateral inhibition between phonemes and words. Inhibition in

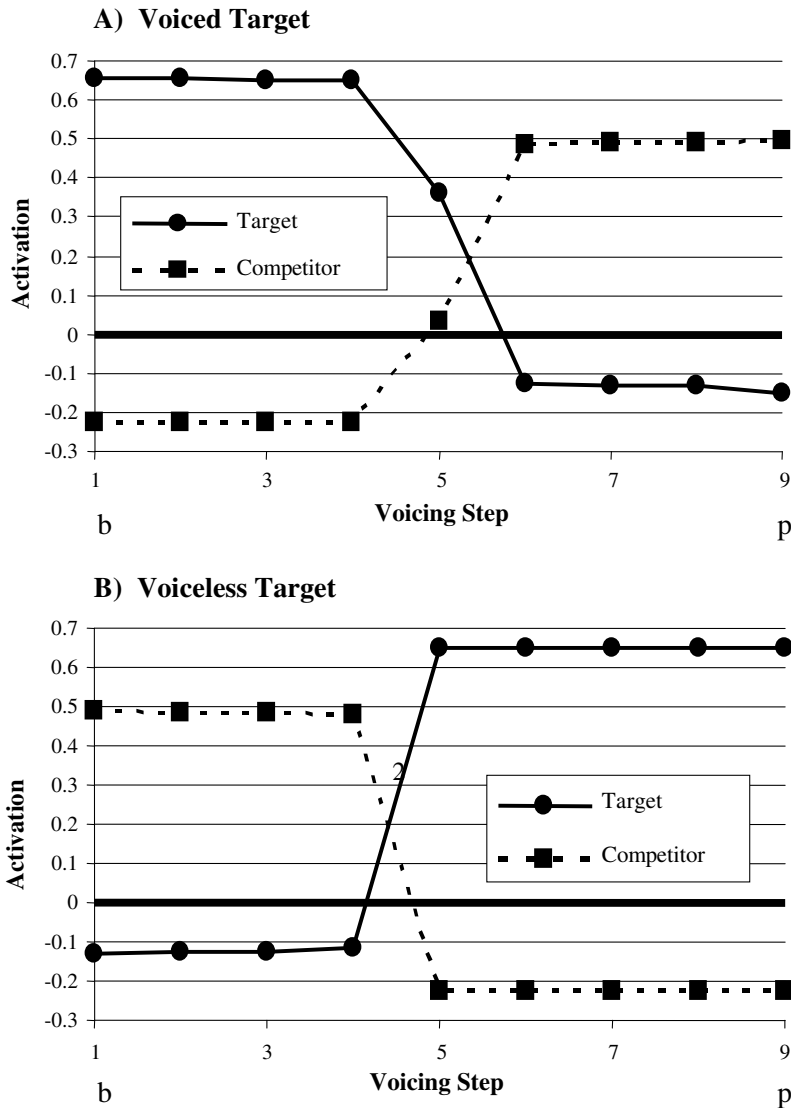


Fig. 5. Final activation (after 200 processing cycles) for a *barricade* → *parricade* continuum (A) or a *barakeet* → *parakeet* continuum (B).

TRACE causes the most active word or phoneme to suppress activation for competitors, moving the system from a state in which multiple candidates (either phonemic or lexical) are active in parallel toward one in which a single item (phoneme or word) has all of the activation. Inhibition therefore seems a likely candidate for why TRACE makes too strong a commitment to a single word, and quickly loses within-category phonetic detail. Panel D shows the results when phoneme inhibition was eliminated. Here, the model provides a good fit for the empirical results, exhibiting both recovery and gradiency. As panel E demonstrates, reducing phoneme inhibition by half is not sufficient to drive this pattern of results—it must be eliminated entirely. Lexical inhibition did not have this effect. Completely eliminating lexical inhibition made the model very unstable (it did not always identify the unambiguous words). However, when the inhibition was halved (panel

F), the model was now able to recover from the garden-path, but its recovery was slow and the influence of stimulus voicing was more or less categorical: steps +0 through +3 showed uniformly quick recovery; +6 through +8 showed equally slow recovery, and +4 and +5 were intermediate. This suggests that continuous variability in voicing may have been lost by earlier processes (in this case phoneme inhibition), but that the lack of commitment at the lexical level allows it to recover nonetheless (from what would be a complete feature mismatch).

The foregoing simulations provide evidence that inhibition seems crucial in both determining whether gradiency can be preserved and whether lexical activation can recover from mismatching input. However, given the flexible dynamics of TRACE, it was possible that similar results could be achieved by other means. Thus, we next assessed the decay parameters. One possibility is that if feature

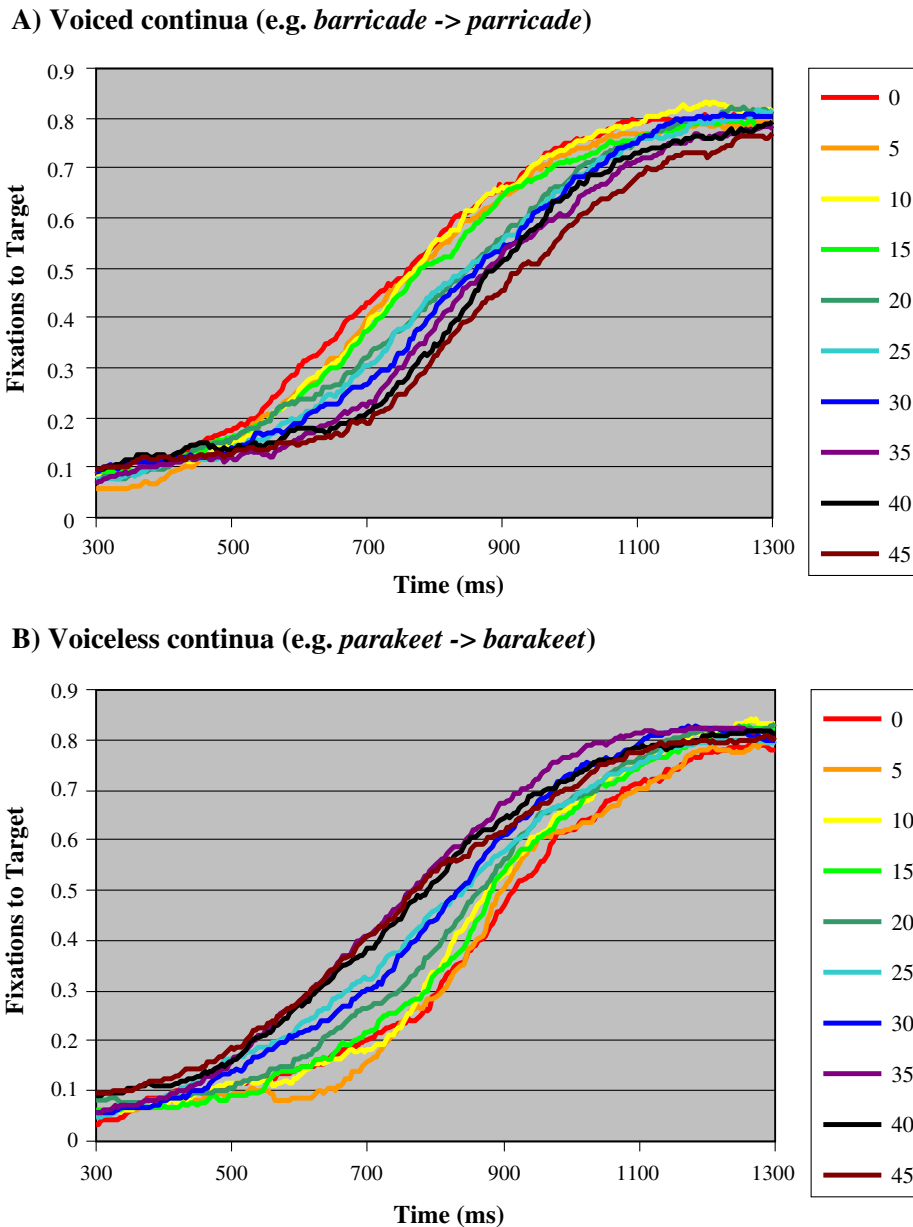


Fig. 6. Proportion of fixations to the target as a function of time and VOT for voiced (A) and voiceless (B) targets in Experiment 1. This data is provided as the closest analog of the simulations reported in Fig. 7.

information decays more slowly, it may be available to facilitate reactivation. However, as panel G shows this was insufficient—the model was still unable to recover from mismatch. We also examined whether slowing the decay rate for lexical and phonetic representations would increase sensitivity to sub-phonetic detail over time. However, doubling the phoneme decay rate (panel H) had no effect. Doubling the lexical decay (panel I) did have the desired effect. However, as in the lexical inhibition simulations, the model's recovery was categorical with respect to voicing—steps +0 through +3 were identical, +6 through +8 were equally slow and +4 and +5 were in between. Once

again, gradience was lost because of earlier processes like phoneme inhibition.

If increasing the decay of lexical activation can help the model recover from mismatch, we hypothesized that weakening its rate of growth might have similar effects. To test this hypothesis we halved the strength of the feed-forward connections between features, phonemes and words. Weakening the feature-to-phoneme connections had little effect (panel J). However, halving the phoneme-to-word weights (panel K) permitted the model to recover from mismatch for at least some of the continua (*billboard*, *pillbox* and *peachpit* but not *beachball*, *barricade* or *para-*

Table 6

The complete set of parameters used in the TRACE simulations, including their default and manipulated values

Parameter	Default	Variant	Panel (Fig. 11)	Description
<i>Activation flow</i>				
Input → Feature	1	—		
Feature → Phoneme	.02	.01	I	½ Feature → phoneme
Phoneme → Word	.05	.1	J	Phoneme → word × 2
Word → Phoneme	.03	0	B	No feedback
<i>Lateral inhibition</i>				
Feature inhibition	.04	—		
Phoneme inhibition	.04	0	C	No phoneme inhibition
		.02	D	½ Phoneme inhibition
Lexical Inhibition	.03	.015	E	½ Lexical inhibition
<i>Decay</i>				
Feature decay	.01	.005	F	½ Feature decay
Phoneme decay	.03	.06	G	Phoneme decay × 2
Lexical decay	.05	.1	H	Lexical decay × 2
<i>Not manipulated</i>				
Feature rest. level	-.1	—		
Phoneme rest. level	-.1	—		
Word rest. Level	-.01	—		
Input noise	0	—		
Feature spread	6	—		
Min. activation	-.3	—		
Max. activation	1	—		

keet). However, this recovery was delayed, and largely categorical.

In sum, across 10 parametric variants, TRACE was largely unable to account for recovery from initial mismatch. The inability to recover from mismatch observed here stands in contrast to now classic (and much debated) simulations reported in McClelland and Elman (1986) demonstrating that TRACE successfully recognizes words like *pleasant* from mismatching input like *bleasant*. The discrepancy between these simulations and ours is most likely due to the lack of competitors for the target word in the limited lexicon of TRACE. Indeed, we were able to replicate McClelland and Elman's (1986) demonstration that the word *rugged* can be recognized given a stimulus of *lugged*. However, there are no *lug*-initial words in the default TRACE lexicon. As *lug*-initial competitors are added to the lexicon, however, TRACE fails to recognize *rugged*. Of course, in a real lexicon, virtually every word will have many cohort competitors—thus, our simulations are more representative of TRACE's behavior with a more realistic lexicon.

To return to our simulations of the garden-path stimuli, however, even when TRACE was able to recover from the mismatch, most of the simulations did not show gradient recovery—that is, TRACE's ability to recover was not influenced by the voicing of the initial syllable. TRACE could only simulate both of these effects when phoneme inhibition was eliminated. The benefits of removing phoneme inhibition are clear when we examine activation of the phoneme units. Fig. 8 shows the raw activation of the /b/ phoneme for a *barricade*/*parricide* continuum as a function of VOT. Panel A shows the default parameters of TRACE. From steps 1 through 4, the model fully activated /b/ with little delay. At step 5, the model does not end up deciding

on /b/ (it chooses /p/), however, activation for /b/ remains sufficiently long enough to be potentially useful. However, by step 6, activation for /b/ decays quite rapidly, and the activation is close to zero when the disambiguating information arrives. In contrast, Panel B shows the same activations when phoneme inhibition is removed. Here the pattern is quite different—from the beginning, both /b/ and /p/ are active and gradiently reflect the voicing in the input. While activation for both gradually decays, this is relatively slow (to the default case) and the gradiency is preserved for both.

In this modified TRACE model, phoneme activation more veridically reflects the input. Both /p/ and /b/ are similarly active because they are both labial stop consonants (the coronal fricative /s/, for example, was not active at this time). The relative difference between them reflects their differences in voicing. This veridical representation of the input then allows lexical competition and feedback to efficiently sort out the input since it is not hindered by an earlier decision. Even when the continuous value of the input is irrelevant (e.g., complete phonemic mismatch), such decisions can be detrimental. For example, /b/ and /p/ share manner and place features, even if voicing is unambiguous. By deciding that a given input is one or the other, the system ignores the possibility that it was incorrect on only one of the three features and is prevented from revising this decision. Thus, minimal sublexical processing may in fact be optimal, provided that phoneme-level inhibition is reduced to prevent early commitment.

Experiment 2

Although the results of Experiment 1 provide compelling evidence for long-lasting gradiency on the way to resolving lexical ambiguity, the robustness of these results would be bolstered by addressing a potential problem with the design. In Experiment 1 the presence of pictures corresponding to the voiced and voiceless competitors in the display might increase the activation of each competitor. Pairs like we used are unlikely to be of equal salience in normal language use. (One is unlikely to talk about *barricades* and *parakeets* in the same conversation.) We were therefore concerned that our results might overestimate the extent to which listeners can recover from lexical garden-paths and the duration over which sub-phonetic detail is available. Experiment 2, addresses this concern in a Visual World study by not displaying the competitor with the target. In Experiment 3 we go beyond the Visual World paradigm entirely by conducting an auditory lexical decision studies designed to further evaluate the possibility that listeners are unusually good at recovering from lexical garden-paths in Experiments 1 and 2 because of repeatedly presenting a small set of stimuli and in the context of visual referents.

Methods

Participants

Twenty University of Rochester undergraduates served as participants in this experiment. All reported normal hearing and normal or corrected-to-normal vision. In-

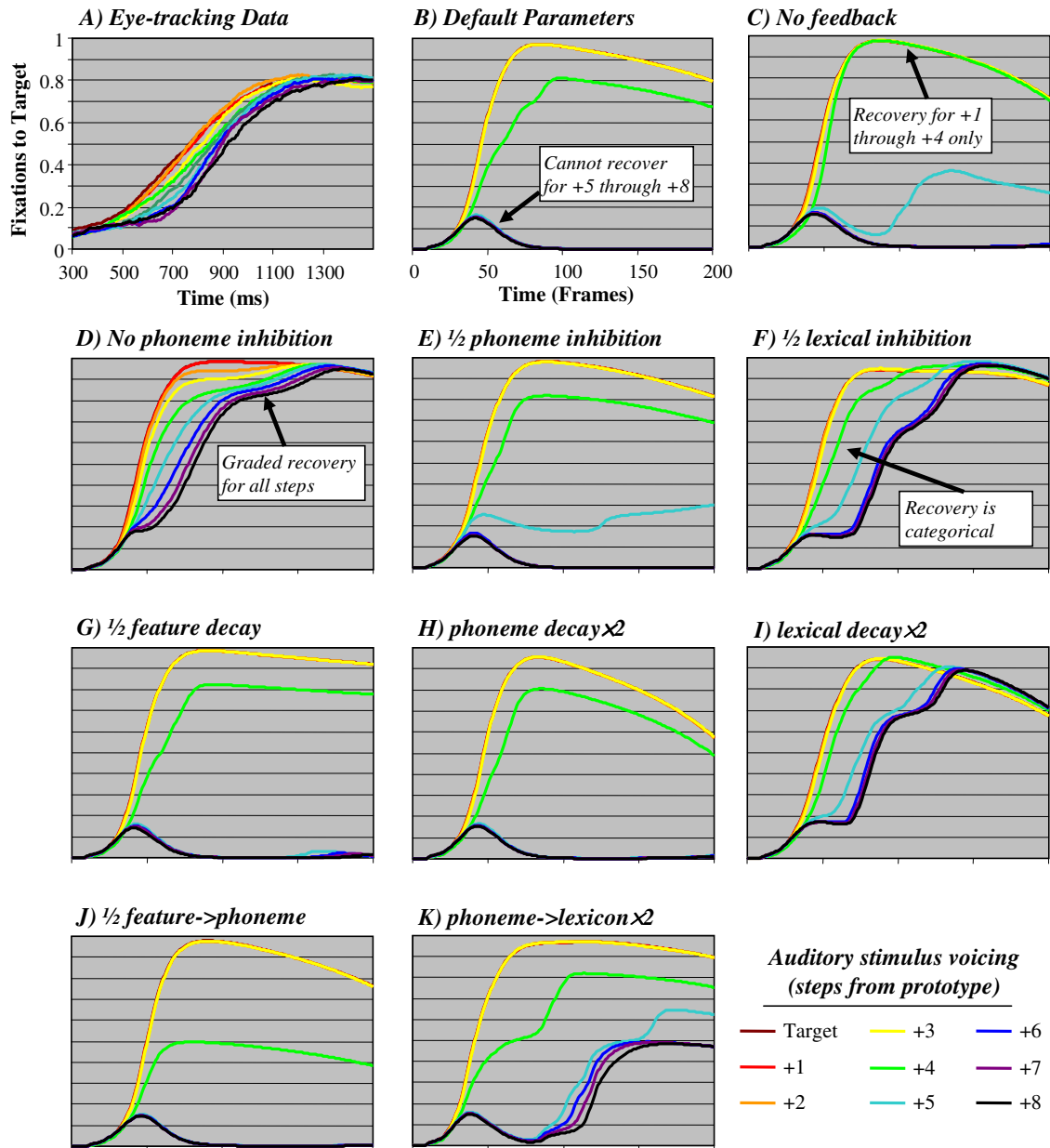


Fig. 7. Predicted fixations as a function of time for TRACE simulations of Experiment 1. Each panel represents a single variant of one parameter. All panels have identical X and Y axes—labels were left off for visual clarity. See Appendix A for the specific parameter values.

formed consent was obtained in accordance with University and APA ethical guidelines. Testing was conducted in two sessions lasting approximately 1 h and participants were compensated \$10/session. One participant was excluded from analysis for failure to return for the second session of testing.

Stimuli

Experimental materials consisted of the same 10 pairs of phonemically similar words (Table 1) that were used in Experiment 1 (10-step continua), along with the same visual stimuli.

Procedure

The task used in Experiment 2 was the same as Experiment 1, with the exception of the particular set of pictures visible on the screen. As before there was no deadline for responses and participants responded in about a second and half ($M = 1560$ ms, $SD = 158.9$ ms).

Design

Experiment 2 used a different arrangement of pictures than Experiment 1. Specifically, the competitor (e.g., *parakeet* for *barricade* trials) was not visible at the same time as the target. However, because participants were likely to be

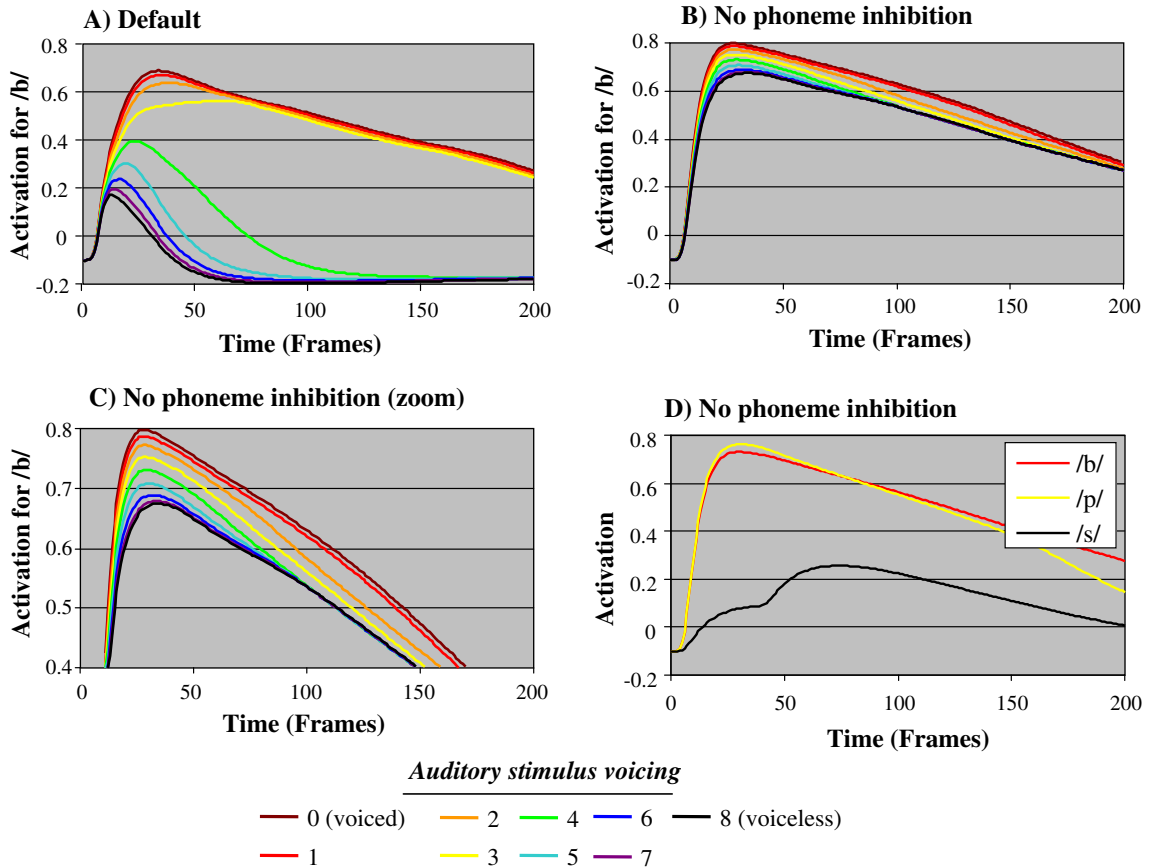


Fig. 8. Activation in the phoneme units during recognition of a *barricade* → *parricade* continuum. (A) Activation of /b/ as a function of time and VOT for default parameters. By step 5, /b/ is completely inactive. (B) Activation of /b/ as a function of time and VOT for simulations without phoneme inhibition—/b/ is active across all VOTs. (C) Zoomed-in view of simulations reported in (B) demonstrating that activation for /b/ is a gradient function of VOT. (D) Activation for /b/, /p/ and a filler, /s/ as a function of time for simulations without phoneme inhibition. While /b/ and /p/ are both active both are much more active than an unrelated phoneme.

at ceiling in identifying a target from among three completely unrelated objects, it may not have been possible to see any effects of the non-displayed competitor in the patterns of fixations to the target. Thus, the competitor object was replaced with one of the other target objects that shared the initial phoneme (e.g., *barricade* and *bassinets*). When this non-target word was presented in the context of the display containing the target object, it served in effect as a very short-overlap cohort competitor. This technique has been used in other studies that have examined possible effects of non-displayed competitors on target fixations (e.g., Dahan et al., 2001b). Similar to Experiment 1, the target was consistently paired with this competitor (as well as two fillers) throughout the experiment, although this set of four pictures was randomized between participants and pairings were selected to avoid any semantic overlap.

To achieve this design, four voiced words (e.g., *barricade*) were chosen along with their voiceless counterparts (e.g., *parakeet*) to create a list of eight possible targets. This list was selected to include no more than two of the t/d pairs and no more than four of the b/p pairs so that there would be enough remaining items to select competitors

that matched on the same phoneme (not just the same voicing condition). Next competitors were chosen to match the first phonemes of the eight targets selected. Note that these selections were made independently of each other. For example, while *barricade* may have *beachball* as a competitor, *parakeet* was not guaranteed to have *beachball*'s counterpart, *peachpit* (and it rarely did). Since on any given trial either of the two items could serve as the target (in this example, *beachball* or *barricade*), this resulted in each participant being exposed to eight sets of visual items, but 16 different continua.

Given this slightly larger number of continua compared to Experiment 1, the number of repetitions at each step was reduced from seven to five. We used the same 10-step continua from Experiment 1, resulting in 800 experimental trials. An additional 640 filler trials were included, yielding 1440 trials, which were run in two sessions lasting approximately an hour.

Results

The analysis of Experiment 2 was similar to that of Experiment 1. We first asked whether VOT influenced the

likelihood that participants would choose the target picture rather than the X, when the onset of the auditory stimulus was mismatching. We then examined the time to fixate the target when the initial fixation was on one of the competitors and the first fixation after the point of disambiguation was to the target. In contrast to Experiment 1, however, we did not examine the probability of fixating the competitor object prior to the point of disambiguation to establish that participants were making an initial graded commitment (similar to McMurray et al., 2002). This was not possible in Experiment 2 because there was no competitor present on the screen to fixate.

Mouse-click evidence for gradient recovery

Participants chose the target that was consistent with the final segment 91% of the time for voiced targets and 85% of the time for voiceless targets. When non-word (X) responses were excluded these figures increased to 99.3% and 98.9%, respectively, demonstrating that non-word responses accounted for the large majority of non-target responses. Interestingly, overall X (non-word) responding was reduced in this experiment (despite identical auditory stimuli) compared to Experiment 1. Recall that we were concerned that the availability of the competitor in the display in Experiment 1 might have inflated effects of within-category phonetic detail and ease of garden-path recovery. Instead, however, participants were more successful in Experiment 2 at using subsequent phonetic context to resolve a mismatching onset segment when the competitor was not displayed. Thus the system seems capable of recovering from mismatching onsets quite gracefully. In fact, the visual competitor may have enhanced competition between the two lexical competitors, leading to the possibility (on some trials) that neither alternative wins, and a non-word response is given.

Fig. 9A shows the proportion of fixations to the target, competitor, and X for the voiced continua (e.g., *barricade/parricade*). As in Experiment 1, the overall pattern of identification results was not consistent with a sharp categorical boundary. The proportion of non-word choices increased gradually as a function of VOT. In addition, as in Experiment 1, even at the most extreme mismatch (*parricade* with a VOT of 45 ms), participants chose the target 73.3% of the time ($SD = 26.2\%$) and only 3 of the 20 participants responded with the X more than the target. For voiceless continua (Fig. 9B), participants responded with the target significantly more than the X at all VOTs, and only 5 out of 20 responded with more Xs than targets at the most extreme mismatch (*barakeet* with a 0-ms VOT).

Eye-movement evidence for gradient recovery

The next series of analyses examined the sequence of fixations to determine if recovery from the garden-path was affected by VOT. Since the competitor belonged to the same voicing category as the target, it did not represent a garden-path interpretation. Thus, the present analysis adopted the more relaxed criteria used in the second analysis of Experiment 1. Trials were included in these analyses if (1) the participant was fixating any object other than the target prior to the POD and (2) the next fixation after the POD was directed to the target. For each trial, the time be-

tween the POD and this first fixation was coded as the time-to-target. Mean time-to-target as a function of VOT is shown in Fig. 10.

These data were analyzed in a 2 (lexical-endpoint) \times 10 (VOT) ANOVA (Table 7). One participant was excluded for missing data. The ANOVA yielded a non-significant main effect of lexical-endpoint by both participants and items (row 1). More importantly the main effect of rVOT was significant, (row 2), as was the linear trend (row 3). Participants were faster to fixate the correct target for VOTs near the prototypical value (i.e., away from the category boundary). This did not interact with target-type (row 4). Thus even with no visible competitor on the screen, variation in VOT was systematically related to the time it took to correctly identify the target after the POD.

The foregoing analysis excluded one participant (who was missing a data-point at a single VOT), and therefore was repeated as a hierarchical regression analysis, similar to the one conducted in Experiment 1 (see Table 8). Again participant codes were added in the first step and found to account for 42.5% of the variance. In the second step, rVOT and lexical-endpoint accounted for an additional 2.1% of the variance, but only VOT reached significance individually ($B = .81$ ms/VOT, 95% $CI_{\text{slope}} = .43$). Finally, the interaction term was added and accounted for no additional variance.

Results were the same when this analysis was repeated with items as the fixed effect. On the first step item codes accounted for 50.6% of the variance. When VOT and lexical-endpoint were added in the second step, an additional 7.6% of the variance was accounted for, and both factors were individually significant (Endpoint: $B = 9.84$ ms, 95% $CI_{\text{slope}} = 8.0$; VOT: $B = .75$ ms/VOT, 95% $CI_{\text{slope}} = .48$). Finally, the interaction term did not significantly account for any additional variance on the third step.

Categorical or gradient?

In order to verify that the linear model relating time-to-target and VOT was superior to a logistic (categorical) model, we again compared two mixed effects models in which linear and logistic functions were applied to each participant's data individually. As before, datasets were constructed for each participant, consisting of the time-to-target for trials in which the participant was fixating any non-target object. On average there were 252.0 data-points per participant ($SD = 75.1$).

Results strongly supported the linear model. The linear model accounted for 9.9% of the variance, and the categorical model for 2.5%. More importantly, the BIC measure showed a difference of 620.1 in favor of the linear model (Linear: 62,440, Logistic: 63,060), and all 20 participants had a lower BIC score for the linear model than the logistic model. Thus, this analysis provides evidence against a categorical model and in favor of a gradient model as the underlying form of the function relating time-to-target to VOT.

Discussion

Experiment 2 replicated the central results from Experiment 1. Recovery from a lexical garden-path was again re-

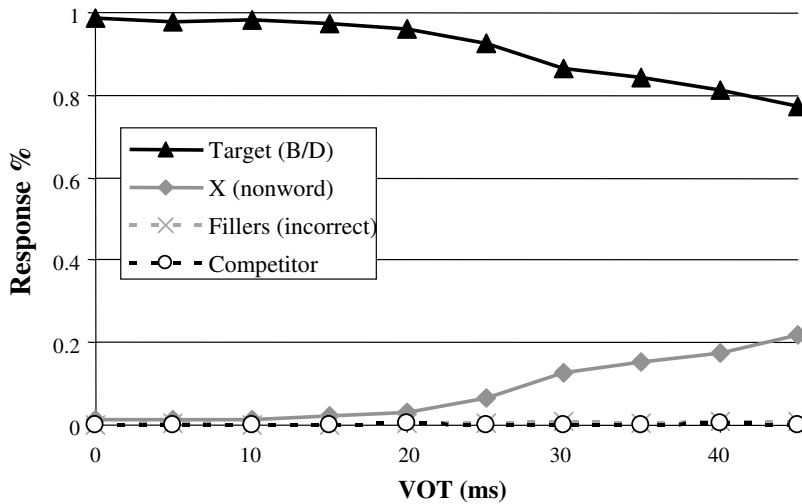
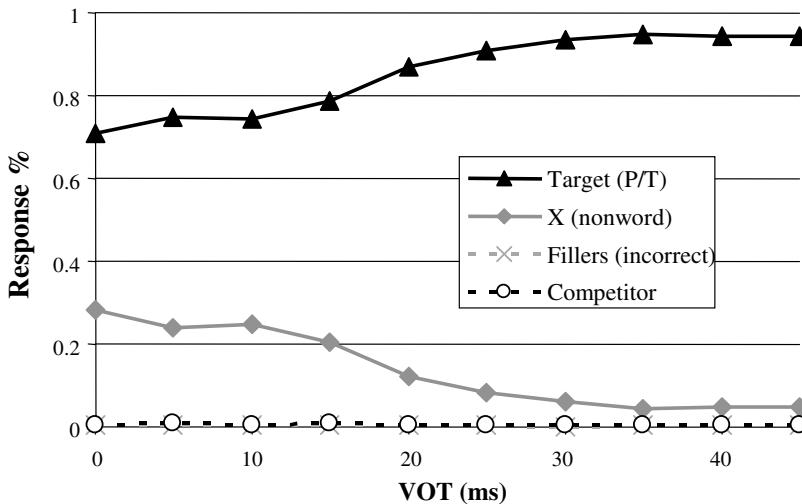
A) Voiced Continua (e.g. barricade->parricade)**B) Voiceless continua (e.g. parakeet->barakeet)**

Fig. 9. Identification results for Experiment 2. (A) Identification for voiced targets (e.g., *barricade*). (B) Identification for voiceless targets (e.g., *parakeet*).

lated to the VOT of the initial consonant, demonstrating that sensitivity to fine-grained sub-phonetic detail persists for at least 240 ms after the POD. The results further demonstrate that gradient effects do not depend on the presence of a visual competitor within the display. Instead the ability of listeners to recover from the mismatch (as seen in their mouse-click responses) was significantly enhanced when the visual competitor was removed in Experiment 2. One likely explanation for this difference is that in Experiment 1, the integration of visual and linguistic information made garden-path recovery more difficult. That is, participants who were looking at the competitor while it was consistent with the input had additional evidence that favored the competitor, whereas participants in Experiment 2 had no picture of the competitor available for fixa-

tion. Thus, there were two sources of information to prevent (or delay) an eye movement to the target in Experiment 1, but only one source (linguistic) in Experiment 2. An alternative is a more strategic explanation, that the presence of the visual competitor simply raised subjects' threshold for what counts as a positive exemplar of the target.

Either way, the fact that across both experiments participants were able to recover from the initial mismatch at all was somewhat surprising in light of classic work suggesting that mismatching word onsets can impede recognition (e.g., Marslen-Wilson & Zwitserlood, 1989). This raises the possibility that either the presence of the visual target or the many repetitions of correctly pronounced targets may artificially raise its activation, and permit recov-

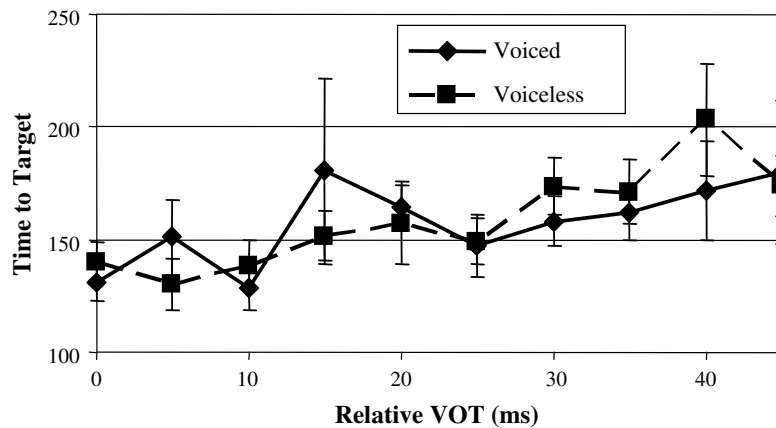


Fig. 10. Time to fixate the target after the point of disambiguation as a function of distance from lexical-endpoint in VOT for voiced and voiceless targets. Time-to-target was computed for trials in which the subject was fixating any non-target object.

Table 7

Results of an ANOVA examining the effect of lexical-endpoint and rVOT on the time-to-target in Experiment 2

Factor	<i>df</i>	<i>F</i>	<i>P</i>	
1 Lexical-endpoint	<i>F</i> ₁	1, 18	3.4	.083
	<i>F</i> ₂	1, 9	3.5	.093
		1, 24	1.7	.2
2 rVOT	<i>F</i> ₁	9, 162	4.8	.0001
	<i>F</i> ₂	9, 81	3.3	.002
	min <i>F</i>	9, 185	2.0	.04
3 rVOT (trend)	<i>F</i> ₁	1, 18	29.8	.0001
	<i>F</i> ₂	1, 9	30.6	.0001
	min <i>F</i>	1, 24	15.1	.0007
4 Endpoint × rVOT	<i>F</i> ₁		<1	
	<i>F</i> ₂		<1	

ery. Thus, in Experiment 3, we eliminated repetition and the visual context by using a lexical decision task.

Experiment 3

Early versions of the cohort model (Marslen-Wilson, 1987) argued that the initial set of activated words is determined by their match to the first one or two phonemes. Thus, auditory input that mismatches a word at onset (e.g., *parricade*) would not enter the cohort and may

never achieve recognition. As a result of this strong theoretical stance, there have been a large number of studies assessing participants' ability to recognize words with mismatching offsets, such as *cathedral* vs. *cathedruke*.

Interestingly, there have been no studies in which participants were asked to make an overt identification response to words that mismatch at onset, such as *shigarette* vs. *cigarette* or *bleasant* vs. *pleasant*. Instead, most have assessed the effect of onset-mismatch on priming. When the onset phoneme mismatches a word by multiple features, this typically prevents priming (e.g., Connine, Blasko, & Titone, 1993; Marslen-Wilson & Zwitserlood, 1989). However, when the onset phoneme mismatches by only 1–2 features, priming can be observed (Connine et al., 1993; Milberg, Blumstein, & Dworetzky, 1988).

However, it is unclear how to interpret these priming data in terms of word recognition. Marslen-Wilson, Moss, and Van Halen (1996), for example, found priming for both close and far mismatches (though there was a numerical, but not statistical difference), but found less priming for mismatches than whole words. They concluded that onset-mismatch degrades recognition. It is not clear, however, if this prevents recognition. Moreover, there is a concern that priming tasks may make participants overly sensitive to mismatch as they must detect non-words on some proportion of the trials. This is buttressed by Frauenfelder, Scholten, and Content (2001) who used a phoneme

Table 8

Results of a hierarchical regression analysis examining the effect of lexical-endpoint and rVOT on the time-to-target in Experiment 2

Analysis	Step	Factor	<i>R</i> ² _{change}	<i>df</i>	<i>F</i> / <i>t</i>	<i>P</i>
Subject analysis	1	Subject	.425	19, 380	14.8	.0001
	2	Main effects	.021	2, 378	7.2	.001
		Lexical-endpoint		378	.9	>.1
		rVOT		378	3.7	.001
	3	Interaction	.0001	1, 377	<1	
Item analysis	1	Items	.506	9, 190	21.6	.0001
	2	Main effects	.076	2, 188	17.13	.0001
		Lexical-endpoint		188	2.4	.016
		rVOT		188	5.3	.0001
	3	Interaction	.006	1, 187	2.9	0.088

decision task to demonstrate lexical activation despite a mismatching onset.

Given this background, it seems clear that the single-feature mismatches used in Experiments 1 and 2 should drive some activation. This is particularly true given the length of the words—there is significant post-onset material available to aid in recovery. However, it has not yet been shown that this would result in word recognition, and it is possible that in Experiments 1 and 2, the presence of the visual target, or the repetitions of the correct and mismatching tokens, created unusual conditions that resulted in complete recognition and exaggerated gradient effects.

The goal of Experiment 3 was to determine if participants are able to recognize the target words at all in the presence of onset-mismatch, when the visual context and the multiple repetitions were removed. The stimuli consisted of both endpoints of the original b/p continua, both the lexical (e.g., *barricade* or *parakeet*) and the mismatching (e.g., *parricide* or *barakeet*) tokens. Intermediate VOTs were not tested. In addition a third list of pure non-words was created by modifying 2–3 additional phonemes of the target items. Finally, each stimulus was only presented once to eliminate the possibility that repeated exposure to the words in Experiments 1 and 2 might have contributed to the low rate of non-word responses.

Methods

Participants

Twenty-five University of Iowa undergraduates served as participants in this experiment. All reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA ethical guidelines. Testing was conducted in a single session lasting approximately 10 min and participants received partial course credit.

Stimuli

Experimental materials consisted of the same 10 pairs of phonemically similar words that were used in Experiment 1. For each word, three conditions were created based on the synthetic continua used in the previous two studies. First, we used the original item (e.g., *barricade*) with its most extreme VOT (either 0 ms for /b/ or 45 ms for /p/). Second, we used a version that differed only in voicing (e.g., *parricide*), again using VOTs of either 0 or 45 ms. Finally, we constructed an unambiguous non-word version from each base word by changing the initial voicing and at least two other phonemes (typically a vowel and consonant). Stimuli are shown in Table 9.

Procedure and design

When participants arrived in the laboratory, informed consent was obtained and the instructions were given. On each trial, participants heard a single stimulus and were instructed to press “w” if it was a word, and “n” if it was a non-word.

There were 10 pairs of words (e.g., *barricade/parakeet*). For each word in the pair, there were three conditions: word, single-feature mismatch, and non-word. This re-

Table 9
Stimuli used in Experiment 3

Base word	Single-feature mismatch	Non-word
Bumpercar	P umpercar	P ampertar
Barricade	P arricade	P arrigoode
Bassinet	P assinet	P oshinet
Blanket	P lanket	P runket
Beachball	P eachball	P ichgall
Billboard	P illboard	P illgoarb
Drain Pipes	T rain Pipes	T raim Bipes
Dreadlocks	T readlocks	T ridrocks
Delaware	T elaware	T ilavare
Delicatessen	T elicatessen	T ilicatefen
Pumpnickel	B umpnickel	B unterdickel
Parakeet	B arakeet	B elakeet
Passenger	B assenger	B ashenker
Plankton	B lankton	B rinkton
Peachpits	B eachpits	B udjpits
Pillbox	B illbox	B illpux
Train Tracks	D rain Tracks	D raim Pracks
Treadmill	D readmill	D ruddbill
Telephone	D elescope	D eleboon
Television	D elevision	D ilemision

Bold characters indicate phonemes that were changed from the base word in order to create mismatch.

sulted in 60 total stimuli and participants heard each stimulus once.

Results and discussion

In the word condition participants correctly identified stimuli as words 75.4% of the time ($SD = 8.7\%$), which seemed somewhat low. However, this was driven by a handful of low-frequency words (*Barricade*: 44%, *Plankton*: 48%, *Delaware*: 52%, *Delicatessen*: 52%) which may have been difficult to identify out of context, and two compound words for which the participants' criteria may not have been clear (*Peachpits*: 44%, *Traintracks*: 44%). Importantly, the single-feature mismatching words were identified as words at a similarly high rate of 63.2%. While this was significantly lower than the word condition ($CI_{\text{difference}} = 6.28\%$), it was also much higher than the non-words which averaged 22.6% ($CI_{\text{difference}} = 4.54\%$). These findings provide strong evidence that participants were willing to accept the mismatching words as lexical targets, even without the support offered by any visual stimuli and without repetition.

Fig. 11 shows a summary of the participants' identification responses across the three experiments. For Experiments 1 and 2, the word condition corresponds to their proportion of correct target responses when the VOT was consistent with the target (e.g., a VOT of 0 ms for *barricade* or 45 ms for *parakeet*). The mismatch condition corresponds to the proportion of correct responses when the VOT was inconsistent (e.g., a VOT of 45 ms for *barricade*). Intermediate steps are not displayed and there is no condition analogous to the non-word condition. Results from Experiment 3 are as previously discussed.

While we must be cautious when comparing across tasks (Experiments 1 and 2 used a word identification task while Experiment 3 used lexical decision) several trends are apparent. First, the comparison between Experiments

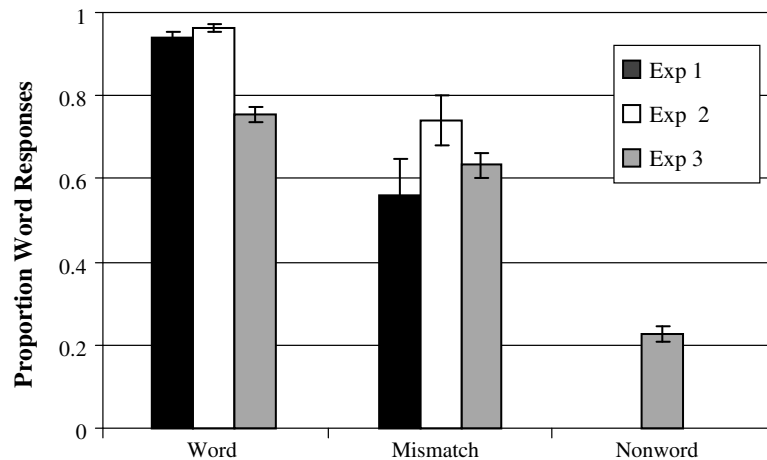


Fig. 11. A comparison of lexical decision responses across the three experiments. Note that in Experiments 1 and 2, the proportion word responses is the proportion of correct target clicks at the endpoints of the VOT continuum. Stimuli analogous to the non-words condition of Experiment 3 were not included in Experiments 1 and 2.

1 and 2 suggests that words that mismatch at onset are more likely to be accepted when the competitor object is not present on the screen (Experiment 2). Thus, our relatively high rate of acceptance cannot be attributed to the presence of the visual competitor. Second, comparing Experiment 2 to Experiment 3 suggests that the visual presence of the target has only a minimal effect on participants' willingness to tolerate mismatch. Thus, Experiment 3 suggests that this finding cannot be attributed to task factors (e.g., the visual objects). Crucially, these results are inconsistent with TRACE, which predicts little ability to recover from onset-mismatch with these stimuli (unless phoneme inhibition is eliminated). Finally, comparing the word conditions in Experiments 2 and 3 it is clear that the presence of the visual target does increase the rate of correctly identifying non-mismatching targets. Presumably the visual context (in Experiments 1 and 2) helps overcome the fact that low-frequency words might be difficult to identify in isolated, decontextualized presentations.

General discussion

The findings reported here make three significant contributions to our understanding of spoken word recognition. First, in contrast to our TRACE simulations, listeners were remarkably flexible in recovering from initial ambiguity. This was particularly apparent in Experiment 2, when the competing word was not supported by the visual context, and in Experiment 3, when there was not any visual context to support the target.

Second, our experimental results provide the strongest evidence to date that gradient sensitivity to small within-category differences in VOT persists over durations longer than a syllable. The evidence comes from participant's interpretation of words such as *parakeet* and *barricade*, which mismatched at onset but otherwise shared the same four initial segments. The likelihood that participants initially fixated the image of the *parakeet* or the *barricade*

was affected by the VOT of the initial consonant, as was the time it took for them to recover from incorrect interpretations. Importantly, this effect obtained throughout the voicing category, and could not be accounted for across subjects by the average of a set of categorical step functions.

Third, our simulations with TRACE demonstrate that phoneme-level inhibition is problematic for models of spoken word recognition. TRACE predicted difficulty in recovering from garden-path errors, and very little effect of within-category VOT on such recovery. Moreover, recovery effects were limited to tokens immediately adjacent to the category boundary and to a brief time-course that did not extend beyond the ambiguous b/p consonant in word-initial position. In contrast, the behavioral results demonstrated that recovery from the garden-path was the norm and that it was systematically sensitive to variation in the first phoneme. Our simulations further demonstrated that phoneme-level inhibition, rather than decay rates or lexical feedback, was the reason why gradient sensitivity to VOT was short-lived in TRACE.

These simulations suggest that any model with strong attractor dynamics at sublexical levels (e.g., phonemes or features) would be likely to underestimate gradient sensitivity to sub-phonetic detail, and would be inflexible in the face of subsequent information. Given the degree of success that TRACE has had in simulating a range of phenomena in spoken word recognition (see Gaskell, 2007, for review), it will be important to determine whether eliminating phoneme inhibition interferes with TRACE's ability to account for these other effects. Evaluating the effects of eliminating phoneme-level inhibition on TRACE is beyond the scope of the present paper (although McMurray, Samelson, Lee, and Tomblin (submitted for publication) demonstrate it has little, if any, effect on the dynamics of activation for cohort and rhymes). What is clear, however, from the simulations we report, is that eliminating phoneme-level inhibition does not make TRACE unstable in the same ways as eliminating word-level inhibition.

Shortlist (Norris, 1994) and MERGE (Norris, McQueen, & Cutler, 2000) do not have inhibition at the phoneme-level. Thus in these models it is possible that sub-phonetic detail could affect garden-path recovery if (a) both alternative lexical items are on the short list at the point of disambiguation⁴ or (b) garden-path recovery occurs in a reanalysis stage that makes rapid use of phoneme-level information, such as the relative activation for the initial phoneme. However, because these models do not attempt to model perception at a sub-phonetic level, it would be premature to conclude that they can better account for the current data than a version of TRACE without phoneme-level inhibition.

The results of our experiments and simulations are also consistent with recent models, based on Bayesian principles, which assume that the likelihood of a category given a probabilistic cue (e.g., the likelihood of a feature, phoneme or a word, given a particular value of VOT) is determined by the variance of that cue, not just its mean or its distance from the mean (see review by Ernst & Bühlhoff, 2004). Clayards, Tanenhaus, Aslin, and Jacobs (in press) have recently demonstrated that listeners do indeed track VOT distributions and behave in accordance with predictions generated by a model based on Bayesian principles (for similar predictions using a Bayesian model of spoken word recognition, see Norris & McQueen, in press). Models of this type have the potential to provide a unified explanation for rapid adaptation and perceptual learning in spoken word recognition (Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003), real-time integration of asynchronous probabilistic cues that participate in trading relations (McMurray, Clayards, Tanenhaus, & Aslin, in press-c), and induction of phonetic categories from distributional information (e.g., Maye, Werker, & Gerken, 2002; McMurray, Aslin, & Toscano, in press-b).

Most generally, regardless of the theoretical stance one takes, it is clear that fine-grained acoustic detail is preserved (in gradient, not discrete form) for at least several hundred milliseconds. In this light, both attractor models (with relaxed sublexical inhibition) and Bayesian models (in which cues are represented as continuous probability values) offer a similar account: during lexical access, the system is sensitive to fine-grained detail, and this detail is retained throughout processing.

Beyond these architectural issues, why might it be important for the word recognition system to maintain fine-grained detail over an extended period of time prior to the point of disambiguation? Since acoustic events are conditioned by the current segment, the preceding and upcoming segments, and the surrounding lexical and sentential context, there are few, if any, points in time at which sufficient perceptual information is available to completely and unambiguously identify the intended word and its component sounds. As a result, the system must continuously deal with the high likelihood that some por-

tion of the signal will be misperceived (or mispronounced by the speaker). Moreover, work in phonetics has argued that variation in cues like VOT may in fact be a source of information toward their underlying causes, and could be used to make inferences about place of articulation, prosodic domain, speaker identity, or even whether a mispronunciation has occurred (e.g., McMurray, Cole & Munson, in press-d; Fougeron & Keating, 1997; Fowler, 1984). In addition, Goldrick & Blumstein (2006) have demonstrated that the value of VOT varies as a function of whether the segment arose from a speech error, an error which would likely be corrected by subsequent sentential context. Several studies suggest that the system can indeed use fine-grained phonetic detail to predict upcoming material (e.g., Gow, 2001; Gow & McMurray, 2007; Martin & Bunnell, 1981). Thus, perception of the current segment can be simultaneously used to build evidence for multiple phonetic events, making the process much more efficient.

The evidence so far suggests that making a sublexical decision of any kind prior to lexical access is at best inefficient and at worst can prevent lexical access. This conclusion would appear to contrast with results from Gaskell, Quinlan, Tamminen, and Cleland (2008) who used the psychological refractory period paradigm to ask whether two sources of phonetic difficulty (coarticulatory mismatch and word/non-word status) affect phoneme decision times before or after a psychological bottleneck. In both cases, they found no evidence for a delay at pre-bottleneck stimulus onset asynchronies, suggesting that their effects occur prior to overt decisional processing, and as part of lexical access. While they interpret these results in terms of a categorical decision process, their phoneme decision task does not rule out the possibility that graded detail is retained despite the overt decision. Indeed, McMurray et al. (in press-a) coupled a similar eye-movement measure to an explicit phoneme decision task to demonstrate that phoneme decision shows similar (though slightly reduced) graded dynamics to lexical activation. Thus, the sublexical categorization processes that occur before the response bottleneck in Gaskell et al. (2008) do not necessarily discard continuous detail.

This raises the important question of whether access to fine-grained detail would be lost with more intervening time (e.g., a longer first syllable in *barricade*), or with more intervening information (e.g., phonemes) between the onset and the point of disambiguation. Work on lexical commitment suggests effects of both time and intervening material: some studies demonstrate that additional information forces a lexical commitment, while other studies point to a role for processing time (see Matty, 1997, for a review). Our study raises a similar question with respect to sublexical commitments. There is however a crucial difference between lexical and sublexical hypotheses. In contrast to sublexical hypotheses, lexical hypotheses do seem to compete over time (Dahan & Gaskell, 2007; Dahan et al., 2001a; Dahan et al., 2001b), allowing for increased time (in the form of extra time for processing/competition), or increased information (altering the evidence for one candidate over another) to affect whether or not alternatives are maintained. While it will be important for future work to examine longer-timescales (and varying degrees of intervening infor-

⁴ James McQueen (unpublished commentary on the second author's Nijmegen Lecture, Eye movements and spoken word recognition, Max Planck Institute for Psycholinguistics, December, 1997) argued that Allopenna et al. (1998) were incorrect in claiming that Shortlist would predict rhyme competition because the rhymes used in that study would not have made it onto the short list.

mation), our study suggests that sublexical representations do not compete to the same degree as lexical representations. That is, the system may not be under pressure to commit to a single sublexical representation and therefore may never make a discrete commitment.

This can be seen clearly, when one considers why the system may be preserving fine-grained detail. Prior work has suggested that a categorical decision regarding initial voicing must be delayed at least until the end of the first vowel (since vowel length is known to contribute to voicing decisions: Miller & Volaitis, 1989; Summerfield, 1981). However, in the present case of words like *parakeet* and *barricade*, such a decision would need to be delayed through 3–5 phonemes before reaching the POD. Since there are no known phonetic contingencies that span such a delay (e.g., there are no further cues to voicing that far from word onset), the fact that continuous aspects of the signal are affecting lexical processes at this late point in the word argues that such a categorical decision is never made. We acknowledge that processes such as between-word inhibition (e.g., Dahan et al., 2001b) and lexical feedback (e.g., Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland, Mirman, & Holt, 2006, for a review) can potentially warp lexical and sub-lexical representations of the input. However, at its most fundamental level, lexical activation is gradiently sensitive to continuous aspects of the signal. This sensitivity, in turn, enables normal activation processes to make use of this detail to perform a rich temporal integration over the signal.

Acknowledgments

This article benefited from helpful suggestions by Gareth Gaskell, Rebecca Treiman and an anonymous reviewer. Special thanks to Arty Samuel, Mary Hare, Jim Magnuson and Delphine Dahan for invaluable feedback on an earlier draft. We would like to thank Dana Subik for assistance with data collection, Dan McEchron for assistance with Experiment 3, and Joyce McDonough for helpful discussions as well as assistance with stimulus presentation. This work was supported by NIH grants DC006537 and DC008089 to BM, and DC005071 to MKT and RNA.

Appendix A. Implementation of the Luce-Choice linking hypothesis

TRACE outputs a set of lexical activations across the entire 220-word lexicon (in this case). However, this must be converted in to the probability of fixating each of the four items on the screen on any given trial (at any given time). We adopt the procedure outlined in Allopenna et al., 1998 which is based on the Luce-Choice rule.

In this procedure, the probability of fixating any given object is given by

$$P(\text{fixating object}_i) = \frac{e^{T a_i}}{\sum_{j=1}^4 e^{T a_j}} \quad (1)$$

Here, a_i refers to the TRACE activation of word i , and T is a temperature parameter (which will be discussed shortly). In a sense, this equation divides the activation (trans-

formed through the exponential function) by the sum of the activations of the four objects on the screen to make a probability.

The temperature parameter controls how veridical this normalization is. At high values of T , the word with the maximal activation tends to assume all of the fixations (e.g., its probability is near one and the others are very low). At low values, the words are more equal. Allopenna et al. discuss the advantages of using a temperature that gradually increases over the course of the trial. Thus, T was defined by

$$T = \frac{10}{1 + e^{(-.05 \cdot (\text{frame} - 90))}} + .5 \quad (2)$$

This equation yielded a logistic function in which T began at a lower asymptote of .5 and smoothly transitioned to an upper asymptote of 10.5, crossing over at frame 90. This dynamic temperature parameter simulates a system that undergoes gradual pressure to settle on a single candidate. However, the findings reported here are not dependent on this assumption—using a fixed T resulted in the same behavior of the model (just slightly poorer fits to the eye-movement data).

Finally, since the Luce-Choice rule given in (1) is a normalized probability, the total fixations at any given time can never be 0 (which is commonly seen in the eye-movement record at the beginning of the trial when subjects are not fixating any of the objects). Thus Allopenna et al. (1998) introduce a scaling factor by which the probabilities computed by (1) are multiplied by the scaling factor in (3) to provide the scaled probabilities.

$$\text{Scaling Factor} = \frac{\max(a_i)}{\max_{\text{overall time}}(a_i)} \quad (3)$$

Here, the numerator is the maximum activation at the current time across the four objects on the screen (target, competitor and the two unrelated objects). The denominator is the overall maximum activation across all four objects over the entire time-course. Thus, by the end of processing the scaling factor is 1 (since the current activation is equal to the final activation), but early in processing it could be quite small.

To sum up, first activations are computed into fixation probabilities using the equation given in (1), where T is defined by (2). Then these probabilities are multiplied by the scaling factor in (3). This is the final predicted fixation probability.

While this introduces a number of free parameters (e.g., the parameters of the logistic describing T , the particular equation of the scaling factors), it is important to note that the overall patterns seen here (e.g., recovery from the garden-path, or failure to recover; gradiency or not) can all be seen in the patterns of raw activation—these free parameters only affect the degree of fit to the eye movements.

Appendix B. The mixed model

The mixed models used to compare gradient (linear) and categorical (logistic) formulations of the data were computed using the following procedure. First, two func-

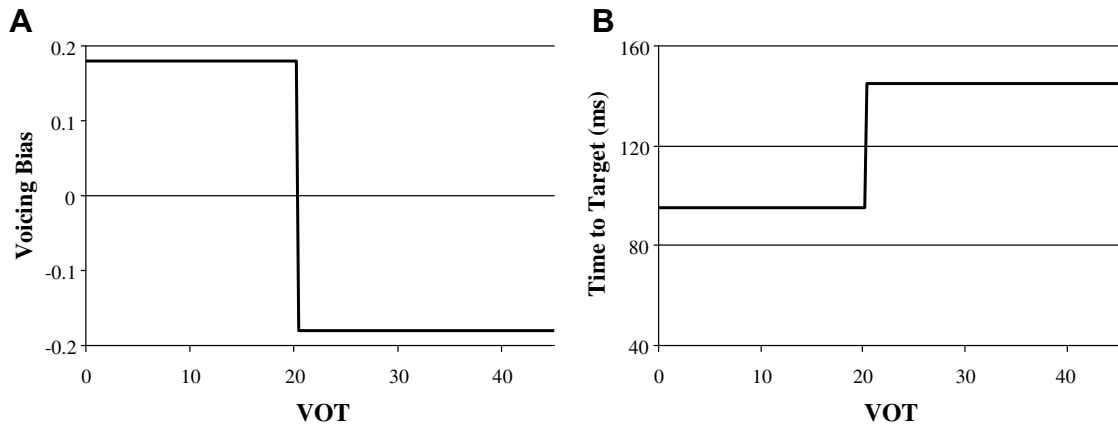


Fig. 12. Categorical step functions used in mixture models. (A) Function used to predict voicing bias for analysis of gradient representations prior to the POD (Experiment 2). (B) Function used to predict time-to-target to examine gradient recovery (both Experiment 2 and Experiment 3).

tions were fit to each subjects' data. The linear function was fit using ordinary linear regression. The step function was based on the logistic function and is given in (4)

$$y = \frac{b_2 - b_1}{1 + \exp\left(\frac{-4s}{b_2 - b_1}(c - \text{VOT})\right)} + b_1 \quad (4)$$

Here, y is either the predicted measure of bias in the initial fixation (for analyses assessing gradiency prior to the POD), or the time-to-target (for analyses assessing recovery from the garden-path). Three variables were free to vary: b_1 , which represents the lower asymptote, b_2 , the upper asymptote and c , the crossover point along the x -axis. The slope of the function is given by s which was set to a constant value of -5000 (for gradiency analyses) and $+5000$ (for Time-To-Target analyses). Fig. 11 shows characteristic functions for predicting initial bias (panel A) or for time-to-target (panel B).

After each function was fit to the data, the log-likelihood of the data given the function was computed. To do this, we assumed that noise was distributed around the function as a Gaussian with some variance, σ , that was estimated from the data directly. The log-likelihood was simply the sum of the natural logs of the likelihood of each datapoint (given the linear or logistic model, and σ).

The log-likelihoods could then be summed across subjects to determine the overall log-likelihood of the mixed model. The Bayesian Information Criterion (BIC) is then computed from this log-likelihood using the equation given below

$$\text{BIC} = -2 \cdot LL + k \cdot \ln(n)$$

Here, LL represents the log-likelihood of the model, and k represents the number of parameters in the model. For the linear model, k was equal to 2 (slope + intercept) \times the number of subjects (17 for Experiment 2, 20 for Experiment 2). For the logistic model, k was equal to 3 (lower/upper asymptotes plus crossover) \times the number of subjects. The variable, n , represents the number of data-points contributing to the model. In the analysis of voiced-bias

prior to the POD in Experiment 2, this was fixed at 100 points per subject (10 VOTs \times 5 continua \times 2 lexical-endpoints). So n was equal to 1700 for this analysis. In the analysis of time-to-targets, this varied between subjects (since the number of qualifying trials was different). Thus, n was 2521 for Experiment 2 ($M = 148.3/\text{subject}$) and 5038 for Experiment 3 ($M = 252.0/\text{subject}$).

Finally, the BIC is not the only criteria used for comparing models with different numbers of parameters. The closely related Akaike Information Criteria (AIC) can compute a similar metric (although it does not take into account sample size). The analyses reported here were also repeated with the AIC, and results were unchanged (in fact the AIC showed an even larger advantage for the gradient model in all cases).

References

- Allen, J. S., Miller, J. L., & De Steno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113, 544–552.
- Alloppenna, P., Magnuson, J. S., & Tanenhaus, M. K. (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.
- Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Blumstein, S., Myers, E., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17, 1353–1366.
- Boersma, B., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.2.31) [Computer program]. Retrieved from: <<http://www.praat.org/>>.
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. (in press). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257–271.
- Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 26, 527–538.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11, 1084–1089.
- Connine, C. M., Blasko, D., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and

- linguistic constraints. *Journal of Memory and Language*, 30, 234–250.
- Connine, C. M., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Dahan, D., & Gaskell, G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57, 438–501.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.
- Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception & Psychophysics*, 62, 843–867.
- Davis, M., Gaskell, G., & Marslen-Wilson, W. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–244.
- Ernst, M., & Bühlhoff, H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162–169.
- Ferrero, F. E., Pelamatti, G. M., & Vaggies, K. (1982). Continuous and categorical perception of a fricative–affricate continuum. *Journal of Phonetics*, 10, 231–244.
- Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3740–36728.
- Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, 36, 359–368.
- Frauenfelder, U., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16, 583–607.
- Gaskell, M. G. (2007). Statistical and connectionist models of speech perception and word recognition. In M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.
- Gaskell, M. G., Quinlan, P. T., Tamminen, J., & Cleland, A. A. (2008). The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*, 137, 282–302.
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649–683.
- Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–139.
- Gow, D., & Gordon, P. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–359.
- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. S. Cole & J. Hualdo (Eds.), *Papers in laboratory phonology* (pp. 173–200). New York: Mouton de Gruyter. Vol. 9.
- Klatt, D. (1980). Software for a cascade/parallel synthesizer. *Journal of the Acoustical Society of America*, 67, 971–995.
- Kopp, J. (1969). A new test for categorical perception. *Psychological Record*, 19, 573–578.
- Kraljic, T., & Samuel, A. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141–178.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107.
- Larkey, L., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, 23, 299–312.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Lieberman, A. M., Harris, K. S., Kinney, J., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, 61, 379–388.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, 60, 484–490.
- Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmas past. *Cognitive Science*, 27, 285–298.
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W., Moss, H. E., & Van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376–1392.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.
- Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, 69, 559–567.
- Matty, S. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, 4, 310–329.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, 101–111.
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10, 363–369.
- McLennan, C., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 539–553.
- McMurray, B. (in preparation). *KlattWorks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research*.
- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (in press-a). Gradient sensitivity to within-category variation in voice-onset time: Effects of stimulus and task. *Journal of Experimental Psychology: Human Perception and Performance*.
- McMurray, B., Aslin, R. N., & Toscano, J. (in press-b). Statistical learning of phonetic categories: Computational insights and limitations. *Developmental Science*.
- McMurray, B., Clayards, M., Tanenhaus, M. K., & Aslin, R. N. (in press-c). Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*.
- McMurray, B., Cole, J. S., & Munson, C. (in press-d). Features as an emergent product of perceptual parsing: Evidence from vowel-to-vowel coarticulation. In N. Clements & R. Ridouane (Eds.), *Where do Features Come from?*
- McMurray, B., Horst, J., Toscano, J., & Samuelson, L. (in press-e). *Connectionism and dynamic systems theory reconsidered*. London: Oxford University Press.
- McMurray, B., Samelson, V., Lee, S., & Tomblin, J. B. (submitted for publication). Eye-movements reveal the time-course of online spoken word recognition language impaired and normal adolescents.
- McMurray, B., & Spivey, M. J. (1999). The categorical perception of consonants: The interaction of learning and processing. *Proceedings of the Chicago Linguistics Society*, 35, 205–219.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86, B33–B42.
- McMurray, B., Tanenhaus, M., Aslin, R., & Spivey, M. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface. Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, 32, 77–97.
- McQueen, J., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1363–1389.
- Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, 26, 305–308.
- Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43, 106–115.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505–512.

- Newman, R., Clouse, S., & Burnham, J. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, 109, 1181–1196.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- Norris, D., & McQueen, J. (in press). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*.
- Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Science*, 23, 299–370.
- Norris, D., McQueen, J., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.
- Salverda, A. P., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.
- Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, 105, 466–476.
- Samuel, A. (unpublished). Perception delayed is perception refined: Retroactive context effects in speech perception.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplement and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, Instruments and Computers*, 39, 19–30.
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of the Acoustical Society of America*, 7, 1074–1095.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, 62, 1297–1311.
- Whalen, D. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, 50, 351–360.