Published in final edited form as:
J Exp Psychol Hum Percept Perform. 2008 December ; 34(6): 1609–1631. doi:10.1037/a0011747.

Gradient sensitivity to within-category variation in words and syllables

Bob McMurray,
Department of Psychology, University of Iowa

Richard N. Aslin,
Department of Brain and Cognitive Sciences, University of Rochester

Michael K. Tanenhaus,
Department of Brain and Cognitive Sciences, University of Rochester

Michael J. Spivey, and
Department of Psychology, Cornell University

Dana Subik
Department of Brain and Cognitive Sciences, University of Rochester

Abstract

Five experiments monitored eye movements in phoneme and lexical identification tasks to examine the effect of within-category sub-phonetic variation on the perception of stop consonants. Experiment 1 demonstrated gradient effects along VOT continua made from natural speech, replicating results with synthetic speech (McMurray, Tanenhaus & Aslin, *Cognition*, 2002). Experiments 2–5 used synthetic VOT continua to examine effects of response alternatives (2 vs. 4), task (lexical vs. phoneme decision), and type of token (word vs. CV). A gradient effect of VOT in at least one half of the continuum was observed in all conditions. These results suggest that during on-line spoken word recognition lexical competitors are activated in proportion to their continuous distance from a category boundary. This gradient processing may allow listeners to anticipate upcoming acoustic/phonetic information in the speech signal and dynamically compensate for acoustic variability.

Keywords

Speech Perception; Categorical Perception; Word Recognition; Subphonemic Sensitivity; Visual World Paradigm

The remarkable ability of human listeners to understand spoken language stands in stark contrast to the highly variable acoustic cues that signal phonemic distinctions. Attempts to find context-independent or invariant cues to simple phonetic features such as place of

Corresponding Author: Bob McMurray, Department of Psychology, E11 SSH, University of Iowa, Iowa City, IA 52242, bob-mcmurray@uiowa.edu, 319-335-2408 (phone), 319-335-0191 (fax).

[9]We were surprised that Experiment 2 shows gradient effects despite the steep labeling function, whereas Experiment 3, with a more continuous labeling function does not. We conducted detailed analyses of the timecourse in order to understand these conflicting results. Two factors appear to conspire to drive these effects. First, Experiment 2 has extremely low variability, making it possible to detect even small differences. Second, our analyses suggest that the, likelihood of detecting gradiency depends on the overall proportion of fixations to competitors. There were relatively more fixations to competitors in Experiment 2 than in Experiment 3, perhaps driven by the fact that there were no other objects on the screen. Details of these analyses are available from the first author.

articulation have met with little success, suggesting that simple invariants do not exist and that speech perception must employ at least partial context-dependent sensitivity. At its heart, the search for invariant acoustic cues in the speech signal is based on the assumption that perception is fundamentally a problem of transforming a continuous and contextually varying signal into discrete and contextually invariant units such as phonemes and words. Continuous aspects of the signal are thought to represent noise that must be discarded during perception, and so-called invariant cues represent nuggets of discreteness buried in this continuous signal.

It is now clear that the variance reduction assumption is unfounded both with respect to the perceptual system, which clearly retains continuity in the signal (Miller & Volaitis, 1989; Andruski, Blumstein & Burton, 1994; Goldinger, 1998; McMurray, Tanenhaus & Aslin, 2002), and with respect to the signal itself, in that few such invariant cues have been found (Lindblom, 1996). The system must balance the efficiency of discarding continuous detail in favor of discrete representations (categorical perception) with the rich information source it offers if it is retained. A long history of phonetic-level work has shown that task variation can alter this balance (e.g. Carney, Widden & Viemeister, 1977; Pisoni & Tash, 1974) in favor of retaining continuous detail. More recently, emerging work shows that word recognition may strongly tilt the scale toward continuity (e.g. Andruski et al, 1994; McMurray et al, 2002; Goldinger, 1998), and that this may benefit online recognition (e.g. Gow, 2001; Gow & McMurray, in press).

The purpose of the present paper is to build on recent work (McMurray et al, 2002) demonstrating that word recognition is systematically sensitive to continuity in the speech signal. First, we ask whether such sensitivity is robust over changes in stimulus (natural vs. synthetic speech). Second, we build an explicit link between the sorts of task variables that have been found to affect categorical perception, and the eye-tracking/word recognition paradigms that have recently shown support for a more gradient speech perception system. To put this in historical perspective, we start by discussing evidence for and against the hypothesis that speech is perceived categorically to illustrate the underlying continuity in speech perception, and the importance of task. We then discuss work showing that this continuity is carried up to higher level processing: spoken word recognition. Finally, we present five experiments that bridge these two domains.

Categorical and Non-Categorical Speech Perception

The variance-reduction view of speech perception was motivated in part by the seminal study of Liberman, Harris, Hoffman and Griffith (1957). They created a synthetic continuum of consonant-vowel (CV) syllables ranging from /ba/ to /da/ to /ga/ by varying the onset frequency of the second formant. In a three-alternative forced-choice labeling task, participants identified exemplars along this continuum as falling into three distinct categories with sharp discontinuities at the category boundaries. Equally importantly, in an ABX discrimination task, participants were only able to discriminate between adjacent exemplars along the continuum if they straddled a category boundary; discrimination of exemplars from within each of the three categories was near chance. Thus, discrimination performance along the continuum was predicted largely by category membership, a pattern of responding that violates Weber's Law, and was named Categorical Perception. This finding has been extensively replicated in both speech (e.g. Ferrero, Pelamatti & Vagges, 1982; Kopp, 1969; Larkey, Wald & Strange, 1978; Liberman, Harris, Kinney & Lane, 1961; Philips, Pellathy, Marantz, Yellin, Wexler, Poeppel, McGinnis & Roberts, 2000; Schouten & Van Hessen, 1992; Sharma & Dorman, 1999) and non-speech domains (e.g. Beale & Keil, 1995; Bornstein & Korda, 1984; Freedman, Riesenhuber, Poggio & Miller, 2001; Howard, Rosen & Broad, 1992; Newell & Bulthoff, 2002; Quinn, 2004). It implies that

continuous detail is lost in favor of discrete, categorical representations, as assumed by variance-reduction approaches to speech perception (see Repp, 1984 for a review).

In contrast to the foregoing idealized view of categorical perception, a large body of evidence suggests that under many testing conditions listeners are sensitive to within-category differences (i.e., *sub-phonetic* detail). In one of the earliest demonstrations, Pisoni and Lazarus (1974) used a 4AIX task in which participants were asked to detect which of two pairs of speech stimuli contained different tokens (the other pair contained identical stimuli). This measure of discrimination revealed that listeners are sensitive to within-category distinctions. Using a same-different task to minimize memory demands, Carney, Widin and Viemeister (1977) also showed evidence of within-category sensitivity. Pisoni and Tash (1974) demonstrated sensitivity to within-category variation in reaction times to single tokens along a synthetic continuum; prototypical exemplars were responded to faster than exemplars near the category boundary. In addition, training studies have shown that category boundaries can undergo changes, that categorical discrimination can be attenuated, and that previously indistinguishable stimuli can be separated into novel categories (Carney et al., 1977; McClelland, Fiez, & McCandliss, 2002; Pisoni, Aslin, Perey & Hennessy, 1982; Samuel, 1977). Finally, a range of studies using both selective adaptation (Miller, Connine, Schermer & Kluender, 1983; Samuel, 1982) and goodness ratings (Allen & Miller, 2001; Massaro & Cohen, 1983; Miller, 1997, Miller & Volaitis, 1989) have shown graded responses within categories. Ratings (or adaptation effects) are highest for prototypical exemplars and fall off monotonically for exemplars near the category boundary, even when all exemplars are judged to be members of the same category.

The foregoing evidence led to a consensus that speech categories are fundamentally graded (as either prototypes or clusters of exemplars) and that these graded categories interact with continuous auditory cues during perception. These interactions can enhance differences between tokens of different categories (e.g. Kuhl, 1991; see Goldstone, Lippa & Shiffrin, 1991 for an example in a non-speech domain). Such an approach does not require a specialized perceptual mechanism or mode; rather this cross-boundary enhancement arises from normal interactive processes during perception (e.g. McClelland & Elman, 1986; Spivey, 2006; McMurray & Spivey, 1999; Anderson, Silverstein, Ritz & Jones, 1977; Damper & Harnad, 2000 for a review).

Importantly, these interactive processes do not eliminate sensitivity to continuous detail, and task and memory demands may alter the relative degree of cross-boundary enhancement and within-category sensitivity. These findings motivate several important questions. First, how do task variables influence this balance between preserving continuous cues and enhancing contrast? The bulk of prior work on the effect of task has examined measures of discrimination (e.g. Carney et al, 1977; Pisoni & Lazarus, 1974; Gerrits & Schouten, 2004; Schouten, Gerrits, E. & Van Hessen, 2003). However, the identification of speech tokens is more fundamental to spoken language processing than discrimination and it is less likely to be confounded with short-term memory issues (Pisoni & Lazarus, 1974; Pisoni, 1973). Second, and perhaps more importantly, does within-category sensitivity to continuous detail affect spoken word recognition? If such information survives early perceptual processes, it may allow the lexical access system to take advantage of systematic regularities in the signal to enhance word recognition processes. In particular, for such information to be helpful, lexical activation must be monotonically related to the continuous cues in the signal, the hallmark of gradient sensitivity.

Gradient Sensitivity in Word Recognition

In the domain of lexical processing, a growing number of studies have found effects of continuous sub-phonetic variation on measures of spoken word recognition. For example,

within-category variation in vowel duration affects the segmentation and activation of embedded words, such as *ham*/*hamster* and *cap*/*captain* (Davis, Marslen-Wilson & Gaskell, 2002; Salverda, Dahan & McQueen. 2003; Salverda, Dahan, Tanenhaus, Masharov, Crosswhite & McDonough, 2007), as well as doubly embedded words, such as *cargo* (Gow & Gordon, 1995). Vowel length also has a continuous influence on the perception of word-final voicing (Warren & Marslen-Wilson, 1988). Moreover, listeners are sensitive to mismatches between coarticulatory cues in the vowel and post-vocalic consonants that can be created through cross-splicing (Dahan & Tanenhaus, 2004; Dahan, Magnuson, Tanenhaus & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 1999; Streeter & Nigro, 1979). However, it is well established that vowels are perceived less categorically than consonants (Fry, Abramson, Eimas & Liberman, 1962; Healy & Repp, 1982). Thus results from vowels may not generalize to consonants. In addition, most of the studies with vowels (a notable exception is Warren & Marslen-Wilson, 1988) were limited to qualitative effects—none included the systematic (and fine-grained) acoustic manipulations that are the hallmark of studies demonstrating categorical perception and that are necessary for evaluating whether there are monotonic effects of within-category variation.

Perhaps the most direct evidence of within-category sensitivity in lexical processing comes from recent studies demonstrating effects of within-category variation for consonants varying in VOT. Andruski, Blumstein, and Burton (1994) used semantic priming to assess lexical activation after participants heard stimuli that were fully voiceless (approximately 80 ms of VOT), 1/3 voiced (53 ms), or 2/3 voiced (27 ms). There was less priming for the target word (e.g., *time*) when it was 2/3 voiced (more /d/-like) than when it was 1/3 voiced or fully voiceless (more /t/-like), suggesting that activation of voiceless targets is reduced as VOT values are shortened to approach the d/t category boundary. Utman, Blumstein and Burton (2000) found the converse effect when primes were the minimal pair competitor (e.g., *dime*, after hearing *time*): there was more priming for the competitor when the target was 2/3 voiced (more /d/-like) than when it was 1/3 voiced or fully voiceless (more /t/-like). These two studies suggest that within-category variation in either the prime or the target affects the magnitude of semantic priming, a measure which clearly taps lexical processing. However, similar to the research on within-category variation with vowels, these studies only assessed a few points along the phonetic dimension of interest, in this case three points on a VOT continuum. The 2/3 voiced stimuli had an average VOT of 27 ms, which is close to the d/t category boundary; there was no effect for the 1/3 voiced stimuli. Thus, one could argue that tokens located near the category boundary (the region of highest uncertainty) were primarily responsible for these effects. Moreover, as a result of the coarse grain of the stimulus manipulation (three 27 ms increments), these studies do not address the question of whether within-category effects are truly gradient (i.e., monotonically related to distance from the category boundary).

A third study of within-category effects on stop consonant perception forms the basis for the present series of experiments. McMurray et al. (2002, see also McMurray, Tanenhaus, Aslin & Spivey, 2003) examined within-category sensitivity using a measure of lexical activation based on the Visual World paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995), as adapted for spoken word recognition (Allopenna, Magnuson and & Tanenhaus, 1998). In McMurray et al. (2002), participants heard tokens from one of six VOT continua instantiated as word-initial consonants (*beach/peach*, *bear/pear*, *bale/pail*, *bomb/palm*, *bump/pump*, *butter/putter*) ranging from 0 to 40 ms in 5 ms steps. After hearing each token, participants selected (with a mouse click) the corresponding picture from a computer display containing the target (e.g., *beach*), its voicing competitor (e.g., *peach*), and two unrelated items (e.g., *ladder* and *shark*). Eye movements were monitored throughout the experiment to provide a measure of the probability of fixating each of the four pictures prior

to, during, and after the target word was presented. As expected, McMurray et al. (2002) showed that the probability of fixating the target picture increased as the word unfolded, and the probability of fixating the competitor picture increased, and then decreased over time. More importantly, the magnitude of this transient competitor activation increased monotonically as the target VOT approached the category boundary. That is, as VOT approached the category boundary participants were more likely to look at the competitor picture—even when the trials on which participants clicked on the competitor were excluded from analysis. These gradient effects were found on both sides of the category boundary. Moreover, linear trends within both phonetic categories were reliable even when data from tokens immediately adjacent to the category boundary were removed. Thus, the evidence for gradiency was not simply the result of uncertainty near the category boundary.

The present series of studies use the Visual World paradigm to provide further evidence that on-line spoken word recognition shows gradient sensitivity to within-category variation in VOT. Although most of the applications of the Visual World eye-tracking paradigm have focused on issues of time course, the paradigm is also well-suited to investigating questions about gradiency. Eye movements are extremely sensitive to fine-grained acoustic variation. But most importantly, the paradigm allows for trial-by trial response-contingent analyses. These analyses reduce the likelihood that what appear to be gradient effects might emerge due to noise in the system. Such noise might arise from the fact that as VOT gets closer to the category boundary, the probability of the stimulus being heard as an exemplar of the cross-category competitor increases. The Visual World Paradigm allows us to filter out such trials. For example, in examining a p/b VOT continuum for *peach/beach*, looks to the picture of a *peach* can be examined for only those trials on which participants indicate that they have heard the input as the word *beach*. This allows an assessment of within-category sensitivity *during* lexical or phoneme identification tasks, something not possible with other measures.

Given standard linking hypotheses[3], the proportion of fixations to competitor words during on-line recognition of the target word reflects the degree of lexical activation of, or evidence for, the competitor (Allopenna et al., 1998;Dahan et al., 2001a;2001b;Magnuson et al., 2003;Tanenhaus, Magnuson, Dahan & Chambers, 2000). Crucially, fixations are sensitive to word frequency, cohort density and neighborhood density, establishing that the eye-tracking paradigm taps into lexical processing (Magnuson, Dixon, Tanenhaus & Aslin, 2007).

We use the same method as McMurray et al. (2002) to monitor eye movements to potential referents during on-line word recognition. First, we replicate the presence of gradiency for natural speech tokens that differ in voicing (e.g., *beach* and *peach*), thereby showing that gradiency is not limited to the use of synthetic speech. We then use this eye-tracking method to ask whether gradiency is present for synthetic speech when the tokens are CV syllables rather than words, and to examine how task variables affect the magnitude of these gradient effects. Even within this relatively simple paradigm, the range of task variables that may play a role is extensive. Here we examined the number of response alternatives, and the word/non-word status of the stimulus, the nature of the task (phoneme decision vs. lexical decision). It was simply not possible to manipulate all of these factors parametrically, which would create more than a dozen experiments. Instead, we selected four representative tasks that sample the space of possible experiments. While this admittedly makes it difficult to isolate specific factors that may change the balance between categorical and gradient representation, it does allow us to identify overall trends.

[3]For example, one such linking hypothesis uses the Luce Choice rule to convert activation across the entire lexicon to fixations to objects on the screen.

Some evidence of gradiency was found in all four of these experiments, demonstrating that phonemic and lexical processing are each sensitive to sub-phonetic details in the input. We argue that this within-category acoustic detail is useful for on-line word recognition, despite the underlying bias to perceive phonetic distinctions categorically using classic phoneme decision tasks.

Experiment 1: Gradiency in natural speech stimuli

The strongest evidence for gradient sensitivity to VOT in lexical access comes from the McMurray et al. (2002) study described earlier. However, a number of studies have suggested that some of the effects documented with single-cue variation, as studied in the laboratory with synthetic speech, may not generalize to natural speech stimuli, which have a richer set of correlated cues. For example, Shinn, Blumstein and Jongman (1985) attempted to replicate two context effects on phonetic categorization: the Ganong effect (in which lexical status influences phoneme category judgments; Ganong, 1980) and the effect of vowel length on manner of articulation judgments. When their continua were varied naturally (e.g., multiple dimensions varying simultaneously), both context effects were reduced or disappeared. When only one or two dimensions were varied in isolation, the context effects clearly emerged. While subsequent research (Miller & Wayland, 1993; see also Burton & Blumstein, 1995) showed that the effects would reemerge with the more complex continua when the stimuli were presented in background noise (multi-talker babble), these studies point out that natural variation in speech may contain perceptual redundancies that could diminish the magnitude of some laboratory phenomena. Most importantly for the present work, processing may be more categorical with natural speech than with synthesized speech. Schouten and van Hessen, (1992) assessed discrimination in a place of articulation continuum created using a spectral averaging technique that yielded extremely natural sounding stimuli. Using this continuum, they found nearly perfect categorical perception. Later work subsequently showed that this effect was reduced when less natural stimuli were created with LPC resynthesis and Klatt synthesis (Van Hessen & Schouten, 1999). In light of these findings, it is important to consider whether the gradient within-category effects reported by McMurray et al. (2002) with synthetic speech would also be found with more natural stimuli.

Experiment 1 uses stimuli constructed from natural speech tokens with the progressive cross-splicing method described by Andruski et al. (1994). Progressively longer portions of the onset of a voiced sound are replaced with similar amounts taken from the aspirated period of a voiceless sound. This creates variations in VOT in which multiple additional cues to voicing are also present (e.g., heightened F1 and F0 and aspiration). Moreover, since such stimuli are constructed from recordings of speech produced by a human vocal tract, they sound much more natural than synthetic speech.

Methods

Participants—Twenty-one University of Rochester undergraduates were recruited from a departmental pool for this experiment. Participants for this experiment and each of the subsequent experiments were recruited and informed consent was obtained in accordance with university human participant protocols and APA ethical standards. All participants were monolingual speakers of English and none reported having hearing difficulties. Participants received $10/day for their participation in this experiment.

Speech stimuli—Auditory stimuli consisted of the same word-pairs used in McMurray et al. (2002). There were six b/p-initial minimal pairs (*beach*/*peach*, *bale*/*pail*, *bear*/*pear*, *butter*/*putter*, *bomb*/*palm*, and *bump*/*pump*), six l-initial filler words (*leaf*, *lamp*, *leg*, *lock*, *lip*, *ladder*), and six sh-initial filler words (*shark*, *shell*, *ship*, *sheep*, *shirt*, *shoe*).

The b/p continua were constructed from recordings of natural speech by progressive cross-splicing (similar to Ganong, 1980; Andruski et al., 1994)[1]. Seven to ten productions of each endpoint word were obtained from a male speaker of American English. Tokens were recorded in a quiet room with a Kay Elemetrics CSL 4300B A/D converter at 11025 Hz. From these tokens, a single pair of endpoints was selected for each of the six continua. Endpoints were chosen such that the voiced endpoint had a VOT as close to 0 ms as possible and the voiceless token had a VOT greater than 40 ms. Over and above this criterion, endpoint pairs were also selected to best match on pitch, sound-quality, and formant frequencies. For each continuum step, a portion of the b-initial tokens (progressively larger portions in approximately 5 ms increments) was removed and replaced with corresponding material from the onset of the p-initial tokens (the burst and aspiration portions). Thus the rhyme portion of each stimulus came from the b-initial tokens, whereas the onset came from progressively larger portions of the p-initial tokens. Splice points were selected at approximately 5 ms increments across the voiceless sections of the p-initial tokens and at corresponding locations in the initial voiced sections of the b-initial tokens, to create 9-step continua with roughly 5ms of VOT between steps. All splices were made at zero-crossings of the waveform, which necessarily resulted in continua that were not uniformly spaced. After construction, the VOT of each continuum-step was measured by careful examination of the spectrogram and waveform (see Table 1)[2].

The same speaker who recorded the test items recorded the filler items. Multiple recordings were made for each of the six l-initial and six sh-initial words, and tokens were selected on the basis of overall quality and prosodic match to the experimental materials (the b/p continua). The filler and experimental items were normalized to the same mean RMS amplitude. Each sound file was preceded by 100ms of silence.

Visual stimuli—The pictures corresponding to each word were identical to those used in McMurray et al. (2002) and are available upon request from the first author.

Procedure—The procedure followed the same protocol as McMurray et al. (2002). After arriving at the lab, an EyeLink II head-mounted eye-tracker was calibrated with the standard 9-point calibration procedure. Participants then read the instructions and began the experiment. Because it was difficult to create easily identifiable pictures for two of the items (*bump* and *putter*), participants received two brief blocks of training trials on the first day of testing to familiarize them with the name/picture mapping. On the first training block, participants saw each of the 24 pictures once in isolation with its name printed below it. On the second block, they saw four pictures (b-, p-, l- and sh-initial items) and the printed name in the center of the screen. In this block, participants clicked on the named picture to advance to the next training trial. Clicking on a picture that did not match the word prevented the trial from advancing and the participant was required to click on the matching picture. There were 48 4AFC training trials, with each picture serving as the target item twice.

After training, the test session began. On each test trial, participants saw four pictures on a 22 in. computer monitor accompanied by a small blue circle in the center of the display. After 500ms (during which participants could scan the screen to determine the location of

[1]For intermediate values along the continuum, this technique can also create mismatching cues. For example, a 10ms VOT stimulus would have the louder burst of a/p/, but the shorter VOT of a/b/. Since our goal was to assess gradiency in natural sounding stimuli (in part so that such stimuli could be used in future experiments), this was deemed a necessary imperfection. This points to the fact that there is no good way to simultaneously manipulate multiple cues and obtain an overall natural-sounding stimulus. The best approach, then, is to test hypotheses with stimuli created using more than one method—the approach we have taken here.

[2]Further information about the construction of these stimuli (along with examples) is available at www.psychology.uiowa.edu/faculty/mcmurray/matss_supplement/

each picture), the central blue circle turned red. The participant was instructed to click on the circle. This established that the mouse was at the center of the screen. As a result the participant was usually fixating the center of the screen at the onset of the target word. Immediately after clicking, the circle disappeared and the participant heard one of the 24 words. Their task was to click on the corresponding picture, which ended the trial. Participants were not encouraged to respond rapidly. Rather, they were told to take their time, to perform the task as naturally as possible, and to ignore the eye-tracker to the best of their ability.

Testing took place in two sessions lasting approximately 50 minutes each. There were no training procedures on the second day. Each session consisted of 540 trials: 5 repetitions at each continua step × 9 continua steps × 6 continua with an equal number of filler trials. Every 54 trials, participants performed a calibration task to allow the software to compensate for any slippage of the eye-tracker on the head.

It was possible that the random assignment of filler items across trials would alert participants to the possible match between b- and p- initial items. Thus, across the entire experiment the b- and p-items were paired with a single set of l- and sh- initial fillers to create a consistent set of four words (e.g., *beach, peach, lamp, and ship*). Each of these sets was maintained throughout the experiment, but was randomized between participants (with the restriction that semantically related pictures, e.g., *beach* and *shell*, never appeared in the same set).

Eye Movement Recording and Analysis—Gaze position was recorded at 250 Hz by a head-mounted EyeLink II eye tracker. This system tracks both eyes and compensates for head-position to yield a dataset containing gaze position in screen coordinates at 4 ms increments. The data-stream from the eye-tracker was automatically parsed into events such as saccades, fixations, and blinks using the system's default parameters.

For analysis, the record of parsed events was converted into fixations on the four displayed pictures that lasted from the onset of a saccade to a specific location to the offset of the corresponding fixation. These were then transformed into a 250 Hz record of which object the participant was fixating at any given time by mapping the average position during the fixation onto the screen coordinates of the displayed items. Items on the screen were 200 pixels tall and 200 pixels wide, and these boundaries were extended by 100 pixels to account for errors of calibration and drift in the eye-tracker. This did not result in any overlap between the four regions on the screen containing the pictures (these regions were each separated by 324, 580, and 664 pixels in the vertical, horizontal, and diagonal directions, respectively).

Eye-movements were recorded for 2000 ms after the onset of the auditory stimulus. On 93% of the trials, participants responded within this window (for the remaining 7%, the eye-movement record was truncated at 2000 ms).

Results

Mouse Click Responses—Given the inherent variability of stimuli constructed from natural speech, it is important to determine how participants labeled the stimuli before analyzing the eye-movement data. Participants selected a filler item for a b/p trial on only . 07 percent of the trials and these trials were excluded from further analysis. Figure 1 displays the results obtained from the final identification response provided by each participant's mouse-click, averaged across participants for each of the six continua. Logistic functions were fit to each of the six continua for each participant to yield estimates of category boundary, amplitude, slope and bias (average R^2=.995). While all four parameters

were estimated for each participant, only category boundary will be discussed in this and subsequent analyses as a source of individual differences.

There was much more variability in category boundary due to item (continuum) than due to participant (SD_{item}=3.44, $SD_{participant}$=2.20). Item variability was consistent across participants ($F(5,80)=46.1$, $p<.001$), with *bear* and *beach* having category boundaries at longer VOTs than the other four continua. Nonetheless, identification functions were smooth and relatively steep, validating both our methodology and our edited natural stimuli.

Fixation Proportions—As in McMurray et al. (2002), only trials in which participants selected the correct target were included in the subsequent analyses. This tended to minimize variation in looks to the target—essentially these fixations were at ceiling—making this measure of limited utility. Analyses focused on the proportion of fixations to the pictures of the competitors, the word corresponding to the other endpoint along a given continuum (for example, if the subject selected *beach,* looks to the *peach*).

Evaluating whether there are gradient looks to competitors for natural stimuli is complicated by variability in VOTs across continuum steps in different continua (due to the cross-splicing manipulation and by variability in category boundaries both between participants and items). Accordingly, we present two analyses that treat VOT slightly differently as a within-participant factor. This was done to minimize the likelihood that we would falsely obtain effects specific to one set of decisions about how VOT and category boundaries map onto discrete steps in an ANOVA framework.

Both analyses use the same dependent measure: the average proportion of fixations to the competitor object over a time window from 300 to 2000 ms after trial onset (similar to McMurray et al., 2002). Since each sound file began with 100 ms of silence and it takes approximately 200 ms to plan and launch an eye-movement, this time window reflected the entire fixation record after the onset of the auditory stimulus that could reasonably have affected any eye movements. The competitor was defined as the picture that would be named by the “opposite” endpoint of the continuum. For example, if the participant heard *bale* with a VOT of 5 ms, *pail* was the competitor. Additionally, both analyses excluded fixation data from any trial where the participant clicked on the picture of the object that was on the other side of their own category boundary. As an example, assume that a particular participant’s category boundary for the *beach-peach* continuum is 20 ms. If that participant heard a stimulus with a VOT that was less than 20 ms, and clicked on the *peach*, then fixations for that trial would not be analyzed because the mouse-click response was “incorrect”. This excluded 4.6% of the trials from the final dataset. The purpose of these exclusion criteria is to yield the most conservative estimate of possible gradient effects by eliminating ambiguous tokens and cross-boundary tokens that were misclassified (on one or more trials). Thus, we limited our analyses to fixations directed to competitors of target words when the correct picture of the target word was selected.

There was one important difference between the analyses used by McMurray et al. (2002) and those presented here. McMurray et al. treated VOT as an absolute variable (in ms). In the current analyses, the distance of each token from each participant’s category boundary for each continuum was used as the independent variable by subtracting the category boundary computed from the mouse-click data from the measured VOT of each stimulus. This will be referred to as relative VOT (rVOT). These computations resulted in voiced sounds receiving negative rVOTs while voiceless sounds received positive values. We chose to use relative VOT rather than absolute VOT because the variability in category boundaries for these items, both between participants and across items, was larger with natural speech

continua than with synthetic speech continua (McMurray, 2004; McMurray, Clayards, Aslin & Tanenhaus, 2004)[4].

Importantly, rVOT analyses are more conservative than analyses that ignore the mouse-click category boundary. By using rVOT, variability in category boundary across participants and continua is effectively eliminated as a potential source of gradient effects. This avoids a well-known problem in the analysis of binomial data in which the average of a number of steep logistic functions with different category boundaries is shallower than any of the contributing functions. While rVOT was adopted in this experiment to deal with the high degree of variability between continua, we also adopted this more conservative analysis for subsequent experiments, even when we used synthetic continua, which are less variable.

Analyses using the ANOVA framework required us to group rVOTs together as discrete levels. While in McMurray et al. this could be accomplished simply using the actual VOT, here, the variability in category boundary, coupled with variability between the VOT step sizes across different continua, required more complex groupings. Since any such grouping is in some ways arbitrary and could introduce bias, we present an ANOVA in which stimuli were grouped by step and a regression analysis in which rVOT is treated as a continuous measure[5]. Importantly, both analyses yielded the expected gradient effects. Figure 2 displays the proportion of fixations to the competitor as a function of time for each rVOT. As hypothesized, the proportion of competitor fixations increases as the rVOTs move closer to the category boundary (absolute values close to 0).

The first analysis took the simplest possible approach by ignoring between-continua variability in rVOT step size and using the continuum step number as the independent variable. This was computed relative to each participant's item-specific category boundary to yield approximately 4–5 steps on the voiced side of the continuum and 4–5 steps on the voiceless side. Two repeated measures ANOVAs were conducted, one for each side of the continuum, with the proportion of fixations to the competitor object as the dependent variable. Both analyses yielded a significant effect of relative continuum step (**B:** $F_1(4,80)=18.1$, $\eta_P^2=.48$, $p<0.001$; $F_2(3,15)=8.0$, $\eta_P^2=0.62$, $p=0.002$; **P:** $F_1(4,80)=20.1$, $\eta_P^2=.50$, $p<0.001$; $F_2(3,15)=19.7$, $\eta_P^2=0.8$, $p=0.0001$)[6], as well as a significant linear trend (**B:** $F_1(1,20)=50.8\eta_P^2=.72$, $p<0.001$; **P:** $F_2(1,5)=44.9$, MSE=0.0001, $\eta_P^2=0.89$, $p=0.001$). For both ends of each continuum, the proportion of fixations to the competitor increased as the continuum step approached the category boundary. To verify that these effects were not primarily due to tokens that were immediately adjacent to the category boundary, these two tokens were removed and a second set of analyses was conducted using the remaining data (7 steps rather than 9 steps). These analyses yielded the same results, with significant main effects of continuum step (**B:** $F_1(3,60)=9.1$, $\eta_P^2=.31$, $p<0.001$; $F_2(2,10)=3.3$, $\eta_P^2=0.39$, $p=0.088$; **P:** $F_1(3,60)=4.6$, $\eta_P^2=.19$, $p=0.006$; $F_2(2,10)=8.2$, $\eta_P^2=0.62$, $p=0.008$) and linear trends (**B:** $F_1(1,20)=26.5$, $\eta_P^2=.32$, $p<0.001$; $F_2(1,5)=3.2$, $\eta_P^2=0.39$, $p=0.13$); **P:** $F_1(1,20)=9.7$, $p=0.006$; $F_2(1,5)=7.9$, $\eta_P^2=0.61$, $p=0.037$) [7].

In the second set of analyses, distance from the category boundary was used as a continuous independent variable in a linear regression. Separate analyses were again conducted on each

[4]In a reanalysis of the McMurray et al.(2002) data using rVOT, all of the analysis reported by McMurray et al. showed the same pattern of significance.

[5]A second ANOVA was performed in which rVOTs were binned into 5ms increments and showed identical results.

[6]Different numbers of VOTs were available for participant and item analyses because the category boundary for the *butter/putter* continuum was shifted towards/b/, away from the other boundaries. Likewise, the *bear/pear* boundary was shifted in the opposite direction. This left one fewer step on the each side for this continuum. Since participant analyses averaged across continua, this resulted in a discrepancy.

[7]While item analyses on the voiced side were only marginally significant, this is likely due to the small number of items used here (6)—note the effect size (ηP^2) which is similar for both item and participant analyses (Participants: $\eta P^2=.32$; Items: $\eta P^2=.39$).

side of the category boundary and the proportion of fixations to the competitor object was used as the dependent variable. Regressions were conducted hierarchically with participant codes entered on the first step to account for between-participant variability (**B:** R^2_{change}=.19, $F_{change}(20,527)=6.0$, p<.001; **P:** R^2_{change}=.24, F(20,564)=8.9, p<.001). On the second step, item codes were entered to account for differences between item-specific continua in the proportion of looks to the competitor (**B:** R^2_{change} =.26, $F_{change}(5,522)=48.2$, p<.001; **P:** R^2_{change} =.03, $F_{change}(5,559)=5.1$, p<0.001). On the final step, rVOT was entered as a continuous variable and was found to be significantly related to looks to the competitor (**B:** R^2_{change}=.03, $F_{change}(1,521)=31.162$, p<0.001; **P:** R^2_{change}=.07; $F_{change}(1,558)=55.8$, p<0.001).

As before, a second set of analyses was conducted in which tokens adjacent to the category boundary were excluded. Tokens 5 ms or less from the boundary were excluded, roughly equivalent to one step in the previous analyses. Significant variance was accounted for by participants (**B:** R^2_{change}=.19, F(20,378)=4.4, p<0.001; **P:** R^2_{change}=.29; $F_{change}(20,456)=9.2$, p<0.001) and items (**B:** R^2_{change}=.34, $F_{change}(5,373)=54.2$, p<0.001; **P:** R^2_{change}=.05; F(5,451)=6.5, p<0.001). Most importantly, small but significant effects of rVOT were found on both sides of the category boundary (**B:** R^2_{change}=.01, $F_{change}(1,372)=7.0$, p=0.009; **P:** R^2_{change}=.01, F(1,450)=4.0, p=0.046). Thus, when treating rVOT as a continuous factor, gradient effects of rVOT were revealed even for tokens that were more than one step removed from the category boundary.

Fixation Durations—Finally, we conducted an analysis in which the duration of fixation to the competitor was the dependent variable. This was done for two reasons. First, as a continuous measure, fixation duration has the potential to make a stronger case for underlying gradiency. Second, the first fixation to the competitor was generally fairly early (543 ms post stimulus onset, 343 ms including the oculomotor delay). Thus effects found with this measure were unlikely to arise from post-decision processes.

Here, trials were included in which the participant initiated a fixation to the competitor at any point after 200 ms (post stimulus onset), and the duration of the first fixation was the dependent variable. Since there were a number of participants who did not fixate the competitor in this time-window (they either fixated it too early, or not at all), a regression analysis was used to cope with the missing data. We predicted that the duration of this first fixation to the competitor would be positively related to rVOT, with longer fixation durations when the VOT was close to the category boundary.

In the first step of the analysis, participant codes were added to the model to account for variability between subjects (**B:** R^2=.07, $F_{change}(20, 1030)=3.9$, SE=130, p<.0001; **P:** R^2=.084, $F_{change}(20,967)=4.066$, p=0.0001). When rVOT was added on the second step, this significantly accounted for an additional 1–2% of the variance (**B:** R^2=.017, $F_{change}(1,1029)=19.2$, p<.0001; **P:** R^2=.011, $F_{change}(1,966)=23.41$, p=0.0001)—as rVOT decreased (moved away from the category boundary) the duration of the first competitor fixation decreased.

For a more conservative analysis, rVOTs within 5 ms of the category boundary were excluded. Results were largely unchanged, with participants accounting for around 8% of the variance (**B:** R^2=.087; $F_{change}(20,690)=3.292$, p<.0001; **P:** R^2=.089, $F_{change}(20,700)=3.4$, p=0.001). rVOT was significant for /b/ but only marginal for /p/ (**B:** R^2 =.011, $F_{change}(1,689)=8.44$, p=.004; **P:** R^2_{change}=.004, $F_{change}(1,699)=3.1$, p=0.081).

Thus the duration of the first fixation to the competitor was systematically related to rVOT. As rVOT departed from the category boundary, the duration of this fixation decreased.

Summary—Experiment 1 found gradient effects of within-category distance from the category boundary as measured by fixations to the competitor picture, regardless of whether rVOT is indexed by continuum step or treated as a continuous covariate. These results do not appear to reflect a post-decision process and can be seen in the continuous duration of the first fixation to the competitor. Moreover, these effects hold even when tokens near the category boundary are removed from the analyses.

Discussion

Experiment 1 replicated McMurray et al. (2002) using stimuli constructed from natural speech. Across a range of analyses we found a monotonic relationship between rVOT and fixations to the competitor picture, even though we adopted extremely conservative criteria that were biased against finding gradient effects. In two experiments, then, we have provided clear evidence that fixations to competitor objects, a measure that reflects degree of lexical activation, shows gradient sensitivity to sub-phonetic detail for VOT.

Our strong evidence for gradient effects comes from an experimental paradigm that differs along at least three dimensions from the two-alternative phoneme identification and discrimination tasks with CV stimuli used in most prior work on categorical perception: (a) the task involved fixations to pictured referents, rather than phoneme judgments; (b) the stimuli were words rather than syllables; and (c) there were four rather than two alternatives. In Experiments 2 through 5, we evaluate whether these factors play a role in determining when gradient effects are observed. In these experiments, we monitored eye movements while manipulating the type of task, the number of response alternatives, and the type of stimuli.

In Experiment 2, we use a two-alternative forced choice task with CV stimuli and a phoneme judgment. This task most closely approximates the two-alternative forced choice button-pressing tasks that typically demonstrate categorical perception. Participants clicked on either an image of a P or a B button on the computer screen. The buttons were displayed on a screen to allow us to monitor eye movements as in Experiment 1. Experiment 3 also uses CV stimuli and phoneme judgments, with the addition of syllables beginning with /l/ and /sh/ to create four rather than two alternatives. The same task is used in Experiment 4, but with the synthesized word stimuli from McMurray et al. (2002). Finally, Experiment 5 uses word stimuli in a two-alternative picture identification task, using a design similar to Experiment 2, but with word stimuli and a lexical task. Although these variations do not explore every permutation of the parameter space, they should reveal whether the stimulus, task, or number of alternatives is primarily responsible for whether or not gradient effects are observed in competitor fixations.

For each experiment, we report identification functions from the mouse-click responses and fixations to the competitor picture as a function of rVOT. We present only minimal discussion of the results of each experiment, deferring more detailed discussion until after we have presented the complete set of data from all four experiments. The primary reason for this deferral is that, although the strength of the evidence for gradiency varies across experiments, some evidence for gradiency is found in each of the experiments, with no single factor accounting for these effects. This becomes clear in a combined analysis which shows overall gradient effects and no significant interaction with experiment.

Experiment 2: Two-choice phoneme identification with CVs

This experiment makes use of a two-alternative forced-choice phoneme identification task using non-words (CV syllables). We used a single 9-step (0–40ms) synthetic VOT b/p continuum which was derived from the lexical continua used in McMurray et al. (2002),

specifically from the *bomb/palm* continuum. As in most prior studies of categorical perception, all consonants were followed by the same vowel, /a/.

Methods

Participants—Nineteen University of Rochester undergraduates were recruited from a departmental participant pool for this experiment. None had participated in Experiment 1. All participants were monolingual speakers of English and none reported hearing difficulties. Participants received $7.50 for their participation in this experiment.

Stimuli—Auditory stimuli consisted of tokens from a synthetic 9-step VOT continuum ranging from /ba/ to /pa/. Stimuli were synthesized with the KlattWorks (McMurray, in preparation) interface to the Klatt (1980) synthesizer[8]. All stimuli began with 5 ms of frication (AF) to simulate the release burst. For stimuli with a VOT of 0 (fully voiced), periodic voicing energy (the AV parameter) began concurrently with this frication. To create non-zero VOTs, the onset of voicing was cut back in 5 ms increments and replaced with 60db of aspiration. First, second and third formants rose to vowel-targets for /a/ specified in Ladefoged (1993) and the entire CV had a duration of 450 ms. This procedure resulted in a /ba/ to /pa/ continuum ranging from 0 to 40 ms of VOT in steps of 5 ms.

Procedure—The procedure was similar to Experiment 1. After arriving at the lab, the Eyelink eye-tracker was calibrated with the standard 9-point calibration procedure. The participants then read the instructions and began the experiment. Each testing trial began with a small blue circle at the center of a 22 in. computer screen. After 500 ms, this central blue circle turned red, at which point the participants clicked on it to initiate the trial. Two large buttons labeled “b” and “p” appeared and a randomly selected token from the VOT continuum was presented over Sennheiser HD 570 headphones. The buttons did not change positions throughout the experiment—*b* was always on the left, and *p* was always on the right. At the same time the two response buttons appeared, a single token from the 9-step VOT continuum was randomly selected and played. The trial ended after the participant clicked on one of the buttons. Participants were encouraged to take their time, to perform the task as naturally as possible, and to ignore the eye-tracker they were wearing.

Testing consisted of a single session lasting approximately 30 minutes. Each of the nine continuum steps was presented 24 times to each of the participants. Thus, each session consisted of 216 trials. A drift-correction calibration was administered every 54 trials.

Eye Movement Recording and Analysis—Gaze position was recorded and analyzed with an EyeLink eye-tracker, using the same techniques reported in Experiment 1. As before, the items displayed on the screen (the *b* and *p* icons) were 200×200 pixels in diameter, and boundaries were extended by 100 pixels on each side for scoring purposes. This did not result in any overlap between the icons and their extended regions, with resulting distances of 324 pixels between the *b* and *p* fixation regions.

Results and Discussion

Mouse Click Responses—Figure 3 displays the average labeling data for each step of the b/p continuum. These functions showed an extremely steep slope, with each VOT being categorized with greater than 90% consistency by each subject. Logistic functions were fitted to each participant’s data, with VOT (0–40 ms) as the independent variable and their labeling results as the dependent measure (average R^2 = .991). This yielded an estimate of category boundary, amplitude, slope and bias for each participant. Overall, category

[8]Information about KlattWorks is available at: www.psychology.uiowa.edu/faculty/mcmurray/klattworks

boundaries displayed an extraordinarily small amount of variability (SD = .56 ms, range=16.2 to 18.7 ms).

Fixation Proportions—The analysis of fixations was conducted using similar techniques to Experiment 1. Any trial in which the participant clicked the "wrong" target (from across the category boundary) was excluded, resulting in the elimination of 2.1% of the trials. As before, the proportion of fixations to the competitor object between 300 and 2000 ms (post trial onset) was the dependent variable.

Despite the fact that the synthesized stimuli simplified the grouping of VOTs (since they had a uniform step-size of 5 ms of VOT between stimuli), we elected to continue to use rVOT (rather than absolute VOT) as our dependent measure. This enabled us to take into account the variation in category boundaries between participants, making it a more conservative test of within-category sensitivity. The distance from each participant's category boundary was used as the independent variable (rVOT) and binned in 5ms increments. Since all of the participants' category boundaries were between the same steps (15 and 20ms of VOT), this analysis was functionally equivalent to an analysis based on absolute VOT (though this was not the case in Experiments 3–5 where more variability was seen).

Figure 4 shows the proportion of fixations to the /p/ button as a function of time and rVOT when the target was from the b-side of the VOT continuum. The first pair of analyses assessed the relationship between rVOT and fixations to the competitor button across the entire range of stimuli. On the voiced side of the continuum, the main effect was only marginally significant (**B:** $F(3,54)=2.7$, $\eta_P^2=.13$, $p=0.053$) and the linear trend was not significant (**B:** $F(1,18)=3.0$, $\eta_P^2=.14$, $p=0.1$). However, on the voiceless side, there was a significant main effect (**P:** $F(4,72)=5.9$, $\eta_P^2=.25$, $p=0.001$) and a significant linear trend (**P:** $F(1,18)=9.9$, $\eta_P^2=.36$, $p=0.006$). When the tokens adjacent to the category boundary were removed, results were unchanged. On the voiced side, there was neither a significant main effect (**B:** $F<1$) nor a linear trend (**B:** $F<1$), whereas both were significant on the voiceless side (**P:** $F(3,54)=3.1$, $\eta_P^2=.15$, $p=0.032$; Trend: $F(1,18)=8.8$, $\eta_P^2=.33$, $p=0.008$).

Fixation Durations—As in Experiment 1, a series of regression analyses assessed fixation duration as a continuous dependent variable. In each analysis the duration of the first fixation to the competitor after 200 ms (post stimulus onset) was the dependent variable. The first analysis examined the voiced side of the continuum. Participant codes were added on the first step and accounted for 12.8% of the variance (**B:** $F_{change}(18,260)=2.1$, $p=0.006$). On the second step, rVOT accounted for an additional .7% of the variance and was not significant (**B:** $F_{change}(1,259)=2.1$, $p=0.144$). On the voiceless side, the first analysis found that participant codes accounted for 10.8% of the variance (**P:** $F_{change}(17,411)=2.9$, $p<0.001$), and rVOT significantly accounted for an additional 3.4% of the variance (**P:** $F_{change}(1,410)=16.3$, $p<0.001$)—as rVOT increased, the duration of fixations to the competitor decreased. A final analysis excluded tokens less than 5 ms from the boundary on the voiceless side. Here, participant codes accounted for 14.1% of the variance (**P:** $F_{change}(17,310)=3$, $p=0.0001$) and rVOT was again significantly related to fixation duration, accounting for an additional 3.1% of the variance (**P:** $F_{change}(1,309)=11.7$, $p=0.001$)—longer fixations were seen for rVOTs near the category boundary. Thus, results with the continuous dependent variable largely mirrored those using fixation proportion, with evidence for gradiency on the voiceless side of the continuum.

Summary—The results from Experiment 2 showed two primary differences with those of Experiment 1. First, the eye-movement results showed reduced sensitivity to within-category detail in the 2AFC phoneme task, with clear evidence for gradient effects on the voiceless side of the continuum, but not on the voiced side.

Second, at a qualitative level the labeling function appeared to have a much steeper slope than Experiment 1 and in McMurray et al. (2002). However, given the range of differences between these experiments (number of alternatives, lexical status, number of continua, response type, etc) it may be premature to make strong claims about the slope of the labeling function. Thus, the next experiment examined the number of response alternatives as a factor by retaining the CV stimuli and phoneme decision task, but using four response alternatives.

Experiment 3: Four-choice phoneme identification with CVs

Experiment 3 used the same CV stimuli as in Experiment 2, but in a 4AFC task like that used in Experiment 1 and in McMurray et al. (2002).

Methods

Participants—Twenty-five University of Rochester undergraduates were recruited from a departmental participant pool for this experiment. All participants were monolingual speakers of English and none reported hearing difficulties. None of the participants had served as participants in Experiments 1 and 2. Participants received $7.50 for their participation in this experiment.

Stimuli—Auditory stimuli consisted of the same synthetic speech syllables used in Experiment 2: a 9-step VOT continuum ranging from /ba/ to /pa/. In addition, two unrelated items (/la/ and /sha/ to match Experiment 1) were synthesized for filler trials. The unrelated items had identical vowels as the VOT continuum.

Procedure—Calibration procedures were the same as in Experiments 1 and 2. Each testing trial began with a small blue circle at the center of a 22 in. computer screen. After 500 ms, the circle turned red, at which point the participants clicked on it to initiate the trial. Four large buttons (labeled *b*, *p*, *l* and *sh*) then appeared and one of the tokens from the VOT continuum (or a filler item) was presented over Sennheiser HD 570 headphones. The participant clicked on the corresponding button and the trial ended. Participants were encouraged to take their time, to perform the task as naturally as possible, and to ignore the eye-tracker they were wearing.

The position of the four buttons did not change throughout the experiment. However, participants may have a bias to make eye-movements to nearby buttons (e.g., after hearing /ba/ participants may be more likely to fixate the *p*-button if it was positioned below the target rather than diagonally from it). Therefore, the particular assignment of buttons to positions was randomized between participants so that for some participants the target and competitor buttons were in a vertical relationship (the closest possible arrangement), for some they were horizontal, and for some they were diagonal (the farthest interbutton distance).

Testing consisted of a single session lasting approximately 50 minutes. Each of the nine steps from the b/p continuum was presented 24 times, yielding 216 experimental trials per participant. Additionally, the two filler items were presented 108 times each, yielding a total of 432 trials. Every 54 trials, participants performed a drift-correction that allowed the eye-tracker to compensate for any slippage of the eye-tracker on the head.

Eye Movement Recording and Analysis—Gaze position was recorded and analyzed with the same equipment and techniques as reported in the previous experiment. Buttons were in the same positions (and the same size) as in Experiment 1.

Results and Discussion

Mouse Click Responses—The mouse click responses were quite accurate and only 0.8% of the trials were excluded due to false alarms (selecting a filler item on a b/p trial). Figure 5 displays the average labeling percentage for each step of the b/p continuum, along with the corresponding labeling data from Experiment 2. The function describing the data for Experiment 3 has a much shallower slope than the function from Experiment 2 and is quite similar to the McMurray et al. (2002) results.

These patterns were confirmed by logistic analyses. Logistic functions were fit to each participant's data with VOT (0–40 ms) as the independent variable and their labeling results as the dependent measure. This yielded an estimate of category boundary, amplitude, slope and bias for each participant (average R^2=.990). An independent samples t-test confirmed that the average *slope* in Experiment 3 (M=.82, SD=.18) was shallower than the slope in Experiment 2 (M=.99, SD=.02; t(42)=4.3, p<.001). This suggests that the steeper slope found in Experiment 2 may derive largely from the number of response alternatives.

Fixation Proportions—The analysis of fixations was conducted using the same techniques as the prior experiments. The distance from each participant's category boundary (rVOT) was used as the independent variable and binned in 5 ms increments. All 25 participants had data for rVOT values of −5, −10 and −15 ms. However, because of variability in the category boundary between participants, only 17 participants had data for an rVOT value of −20 ms. Therefore, we opted to analyze only the first 3 steps on the voiced side (although the general pattern of results was the same when all four steps were included). This was not a problem on the voiceless side where data were available for all 25 participants from +5 to +25 ms.

Figure 6 displays the proportion of fixations to the competitor as a function of rVOT and time after target word onset. Here, the effect of rVOT seems to be driven primarily by tokens adjacent to the category boundary (+5 and −5 ms of rVOT). This was confirmed by separate one-way ANOVAs. There was a significant main effect of rVOT on both sides of the category boundary (**B:** $F(2,48)=9.2$, $\eta_P^2=.28$, $p<0.001$; **P:** $F(4,96)=7.6$, $\eta_P^2=.24$, $p<0.001$) when all rVOT steps were included. The linear trends were also significant (**B:** $F(1,24)=15.3$, $\eta_P^2=.39$, $p=0.001$; **P:** $F(1,24)=10.3$, $\eta_P^2=.30$, $p=0.004$). However, when tokens adjacent to the category boundary were removed from the ANOVA, neither main effect was significant (**B:** $F(1,24)=1.7$, $\eta_P^2=.07$, $p=0.2$; **P:** $F<1$)[10].

Fixation Durations—The first hierarchical regression analysis assessed the voiced side of the continuum. In the first step, participant codes significantly accounted for 12.9% of the variance (**B:** $F_{change}(24,358)=2.2$, $p=0.001$). In the second step rVOT was added and (as in Experiment 2) was not significant (**B:** $R^2_{change}=.002$, $F_{change}(1,357)=1$, $p>0.2$). On the voiceless side, participants accounted for an initial 12.2% of the variance (**P:** $F_{change}(24,475)=2.8$, $p=0.0001$), but rVOT did not significantly account for any additional variance (**P:** $R^2_{change}=.003$, $F_{change}(1,474)=1.8$, $p=0.183$).

Summary—Experiments 2 and 3 both support the hypothesis that some gradiency is visible in phoneme decision tasks, although it may be attenuated compared to lexical identification tasks. However, two differences emerged from the relatively straightforward manipulation in number of response alternatives between Experiments 2 and 3. First, the labeling function in the 4AFC task was shallower than the 2AFC task, presenting an explanation for the steep slope seen in Experiment 2. A classic explanation for such

[10]Note that item analyses were not possible since Experiments 2 and 3 did not use the same set of 6 items.

differences is noise in the perception of VOT *prior to categorization*. This would result in VOTs close to the boundary being occasionally misperceived as VOTs on the other side of the boundary and flattening the slope near the boundary, since VOTs further away would be less likely to be misperceived as tokens of the opposite category. This hypothesis, however, seems unlikely, since it is not clear why the number of response alternatives would change the encoding of VOT. Thus, it would appear that the 4AFC task is more sensitive to the continuous underlying representation of VOT and is revealed in the labeling data (as a shallower slope), consistent with the labeling data of Experiment 1 and McMurray, et al (2002).

Second, while both experiments showed gradient results when the entire continuum was considered, when tokens adjacent to the boundary were removed, results differed. In Experiment 2, gradiency persisted on the voiceless side, but not on the voiced side. In Experiment 3, it was not seen on either side. This is difficult to explain. However, our analysis of pattern of fixations over time revealed somewhat different patterns. In both cases, the curves revealed the standard pattern: early in the trial subjects fixated the competitor, and then these were reduced as they identified the target. In Experiment 2, the maximum of this initial portion was higher than in Experiment 3 (likely due to the fact that there was only one competitor). However, it also fell off rather quickly and was highly consistent between subjects. In Experiment 3, the maximum was lower, but it persisted longer and was much more variable between subjects. The higher peak activation coupled with very low variability may have allowed us to see a very small gradient effect that was likely masked in the other experiment.

Experiment 4: Four-choice phoneme identification with words

It is clear that across both Experiments 2 and 3 non-word phoneme decision tasks with CV stimuli result in reduced sensitivity to within-category detail compared to the 4AFC lexical tasks used in Experiment 1 and in McMurray et al. (2002). However, Experiments 2 and 3 differ from Experiment 1 and from McMurray et al. in both the type of stimuli (words vs. CV syllables) and the type of decision (lexical identification vs. phoneme labeling). Experiment 4 bridges these task differences by using a 4AFC phoneme decision task with the original lexical stimuli from McMurray et al. (2002).

Methods

Participants—Nineteen University of Rochester undergraduates who had not participated in any of the previous experiments were recruited from a departmental participant pool for this experiment. All participants were monolingual speakers of English and none reported hearing difficulties. Participants received $7.50 for each of two sessions in this experiment.

Stimuli—Auditory stimuli were the synthetic word stimuli used in McMurray et al. (2002). These consisted of six 9-step VOT continua ranging from /b/ to /p/ in word/word contexts: *beach/peach, bear/pear, bail/pail, bomb/palm, butter/putter, bump/pump*. The same six l- and sh- initial filler words were also used (*lamp, ladder, lock, leg, lip, leaf, shell, sheep, ship, shark, shirt, shoe*). Each b/p word pair had identical release bursts and rising first and second formants which terminated in the vowel targets described in Ladefoged (1990). As in Experiments 2 and 3, VOT was manipulated by cutting back the temporal onset of the amplitude of the voicing parameter in 5 ms steps and replacing it with 60 dB of amplitude of aspiration. This yielded a 9-step (0–40 ms) VOT continuum for each of the labial-initial word pairs. Syllable-final formant transitions and frication parameters were chosen to match as closely as possible to spectrograms of natural speech and to create the most natural sounding stimuli (see Appendix A for KlattWorks scripts).

Filler items (/l/- and /sh/-initial items) were constructed using vowel targets from Ladefoged (1990) and formant trajectories that matched spectrograms of natural speech. All stimuli (targets and fillers) had identical pitch contours (F0). Stimulus durations, however, differed between tokens and were chosen to maximize naturalness. All items contained an initial 100 ms of silence.

Procedure—The procedure was the same as Experiment 3. Each testing trial began with a small blue circle at the center of a 22 in. computer screen. After 500 ms, the circle turned red and participants clicked on it to initiate the trial. Four large buttons (labeled *b, p, l* and *sh*) appeared and one of the tokens from one of the six VOT continua (or a filler item) was presented over Sennheiser HD 570 headphones. The participant clicked on the corresponding button and the trial ended. Participants were encouraged to take their time, to perform the task as naturally as possible, and to ignore the eye-tracker they were wearing. Most importantly, the instructions emphasized that the participant's task was to determine the identity of the first sound, and not the identity of the word ("pay attention to the first sound in each word and determine if it sounds more like a 'b', 'p', 'l' or 'sh'").

As in Experiment 3, the position of the four buttons did not change throughout the experiment. However, the position of the buttons between participants was randomized so that the distance between target and competitor buttons was counterbalanced across participants.

Testing consisted of two sessions lasting approximately 50 minutes each. Each of the 9 continuum-steps was presented 10 times to each of the participants, yielding 540 experimental trials across the six continua. The two filler items were presented 270 times each, yielding a total of 1080 trials. This design was identical to Experiment 1. Every 54 trials, participants performed a drift-correction that allowed the eye-tracker to compensate for any slippage of the eye-tracker on the head.

Eye Movement Recording and Analysis—Gaze position was recorded and analyzed with the same equipment and techniques as reported in the previous two experiments. The four buttons were in the same positions (and the same size) as the pictures in Experiments 1 and 3.

Results and Discussion

Mouse Click Responses—The mouse click responses were quite accurate and only 0.2% of the trials were excluded due to false alarms (selecting a filler item on a b/p trial). As before, logistic functions were fit (average R^2=.960) to each participant's mouse click data for each item (to account for differences between items in terms of category boundary).. These functions did not differ from those in Experiment 3 in either slope ($t(42)=.6$, $p>.1$; note that the Exp. 4 slopes were computed for each of the six continua and averaged) or category boundary ($t(42)=1.1$, $p>.1$) (see Figure 7). Category boundaries also had similar variability across subjects to those in Experiment 3 (SD = 3.5 ms vs. SD=2.3 ms). Due to the number of factors that differ between these experiments, we must be cautious in making comparisons. While we certainly would not want to conclude that there are no differences between the labeling data of Experiment 3 and 4, these analyses confirm that this 4AFC metalinguistic task of judging the identity of the word-initial phoneme (at least from the perspective of labeling functions) did not differ substantially from the same 4AFC task with CV syllables used in Experiment 3.

Fixation Proportions—As in the previous experiments, distance from each participant's category boundary (rVOT) was used as the independent variable and binned in 5ms

increments. All 19 participants had data for rVOT values of −5, −10 and −15 ms, but because of variability in the boundary values between participants, one participant did not have data for the rVOT value of −20 ms. We opted to analyze all four steps on the voiced side by removing one participant from analyses on this side of the continuum; all results were similar when data from that participant were included. On the voiceless side of the continuum, rVOT values were available from all 19 participants (from +5 to +25 ms).

Figure 8 shows the proportion of fixations to the competitor button as a function of time and rVOT. Separate one-way ANOVAs revealed a significant main effect of rVOT on both sides of the continuum when all of the tokens were included (**B:** $F_1(3,51)=10.4$, $\eta_P^2=.38$, $p<0.001$; $F_2(3,15)=34.5$, $\eta_P^2=0.87$, $p=0.0001$; **P:** $F_1(4,72)=12.3$, $\eta_P^2=.41$, $p<0.001$; $F_2(4,20)=15.0$, $\eta_P^2=0.75$, $p=0.0001$). Linear trends were also significant (**B:** $F_1(1,17)=21.7$, $\eta_P^2=.56$, $p<0.001$; $F_2(1,5)=128.6$, $\eta_P^2=0.96$, $p=0.0001$; **P:** $F_1(1,18)=17.9$, $\eta_P^2=.50$, $p<0.001$; $F_2(1,5)=50.1$, $\eta_P^2=0.91$, $p=0.001$). However, when tokens adjacent to the category boundary were removed from the analysis, the main effect of rVOT (**B:** $F_1(2,34)=1.5$, $\eta_P^2=.08$, $p>.2$; $F_2(2,10)=14.9$, $\eta_P^2=0.75$, $p=0.001$), and the linear trend (**B:** $F_1(1,17)=2.4$, $\eta_P^2=.12$, $p>0.1$; $F_2(1,5)=27.8$, $\eta_P^2=0.85$, $p=0.003$) were significant by items but not by participants on the voiced side, but remained significant by both participants and items on the voiceless side (**P:** $F_1(3,54)=5.2$, $\eta_P^2=.23$, $p=0.003$; $F_2(3,15)=7.2$, $\eta_P^2=0.59$, $p=0.003$; Linear Trend: $F_1(1,18)=5.6$, $\eta_P^2=.24$, $p=0.029$; $F_2(1,5)=16.5$, $\eta_P^2=0.77$, $p=0.01$).

Fixation Durations—A set of heirarchical regression analyses examined the effect of rVOT on the duration of the first look to the competitor. Separate analyses were run for the voiced and voiceless side of the continuum. In the first step of the analysis, participant codes were added and were significant (**B:** $R^2=.065$, $F_{change}(18,844)=3.2$, $p=0.0001$; **P:** $R^2=.101$, $F_{change}(18,892)=5.6$, $p=0.001$). rVOT was added on the next step and accounted for an additional 2% of the variance (**B:** $R^2=.022$, $F_{change}(1,843)=20.8$, $p=0.0001$; **P:** $R^2=.024$; $F_{change}(1,891)=24.4$, $p=0.0001$). This effect was driven by the fact that fixations to the competitor were longer as rVOT approached the category boundary.

Results were similar when rVOTs adjacent to the category boundary (within 5 ms) were excluded. On the first step participant codes were significant (**B:** $R^2=.086$, $F_{change}(18,563)=2.9$, $p=0.0001$; **P:** $R^2=.118$, $F_{change}(18,681)=5.1$, $p=0.0001$). On the second step, rVOT was added. On the voiced side it only accounted for an additional .8% of the variance, but was significant (**B:** $F_{change}(1,562)=5.2$, $p=0.022$). On the voiceless side, it accounted for less and was not (**P:** $R^2=.003$, $F_{change}(1,680)=2.3$, $p=0.13$).

Summary—Overall, the results of this experiment did not depart dramatically from Experiments 2 or 3. The labeling results closely resembled those of Experiment 3 (which had shallower slopes than Experiment 2), suggesting that switching from CVs to words in a 4AFC phoneme decision task has little or no effect on the overall pattern of labeling responses. At the finer grained level afforded by the pattern of eye-movements, results generally paralleled those from Experiments 2 and 3. There was within-category sensitivity for both voiced and voiceless sounds, but only when tokens adjacent to the category boundary were included in the analysis. When these tokens were removed from the analysis, we saw weaker evidence for within-category sensitivity, with the effects failing to reach significance on the voiced side.

Experiment 5: Two-choice picture identification with words

Experiments 2, 3, and 4 all used a task (either 2AFC or 4AFC) in which the listener's attention was focused on the initial consonant of either two or four CVs, or four words. In these phoneme decision tasks there was evidence of gradiency, but it was less robust than

for the lexical identification tasks used in Experiment 1 and in McMurray et al. (2002). Experiment 4, which used lexical stimuli in a 4AFC phoneme decision task, showed a tendency towards stronger gradient effects than Experiment 3, which used non-lexical CV stimuli in the same 4AFC phoneme decision task. This raises the possibility that any lexical choice task with lexical stimuli would show stronger gradient effects, even when the decision is simplified by only having two response alternatives. We evaluate this possibility in Experiment 5 by using lexical stimuli in a 2AFC lexical identification task in which the positions of the pictures were fixed such that b-initial items were consistently on the left and p-initial items were consistently on the right (as were the /b/ and /p/ buttons in Experiment 2).

Methods

Participants—Twenty-two University of Rochester undergraduates were recruited from a departmental participant pool for this experiment. None had participated in the previous experiments. All participants were monolingual speakers of English and none reported hearing difficulties. Participants received $7.50 for participation in this experiment.

Stimuli—Auditory stimuli consisted of the same synthetic lexical items used in McMurray et al. (2002) and in Experiment 4. Six 9-step VOT continua, ranging from 0 to 40 ms, were used, (*beach/peach, bear/pear, bale/pail, bump/pump, bomb/palm and butter/putter*).

Procedure—The procedure was designed to include as many of the simplifying properties of Experiment 2 (2AFC with CV syllables), combined with the lexical task of Experiment 1. Each testing trial began with a small blue circle at the center of a 22 in. computer screen. The circle turned red 500 ms later and the participants clicked on it to initiate the trial. Two pictures then appeared. The pictures were the same size and in the same location as the buttons used in Experiment 2. Simultaneously with picture onset, a token from the VOT continuum corresponding to one of the pictures was presented over Sennheiser HD 570 headphones. The participant clicked on the corresponding picture and the trial ended. Participants were encouraged to take their time, to perform the task as naturally as possible, and to ignore the eye-tracker they were wearing.

While the particular pictures displayed on any given trial were random, their position on the screen was not: /b/ pictures were consistently on the left and /p/ pictures on the right. This fixed positioning, coupled with the lack of filler items, meant that Experiment 5 differed from the 2AFC phoneme judgment task used in Experiment 2 in only three ways: the task was picture identification rather than phoneme identification; the stimuli were words rather than non-lexical CVs; and there were six continua rather than one.

Testing consisted of a single session lasting approximately 50 minutes. Each of the 9 continuum-steps was presented 10 times to each of the participants, yielding 540 total trials. Every 54 trials, participants performed a drift-correction that allowed the eye-tracker to compensate for any slippage of the eye-tracker on the head.

Eye Movement Recording and Analysis—Gaze position was recorded and analyzed with the same equipment and techniques as reported in the previous four experiments. Pictures were in the same positions (and the same size) as the buttons in Experiment 2.

Results and Discussion

Mouse Click Responses—Logistic functions were fit to each participant's identification functions for each of the six continua (average R^2=.968). These functions were generally similar to Experiment 4 (using the same stimuli). The category boundary in Experiment 5

(2AFC words: M=18.3 ms) did not differ from Experiment 4 (4AFC phoneme decision in words; t(36)=1.6, p>.1) (see Figure 10).

More importantly, slopes of the identification functions in Experiment 5 differed from the results of Experiment 2, the other 2AFC task in this series of experiments; identification functions were shallower in the current experiment (M=.70, SD=.16; t(39)=7.8, p<.0001). Note that this difference is not due to variability in category boundary between continua; this variability was eliminated by computing separate slopes for each continuum (within-participant). Slopes did not differ between Experiments 4 and 5 (t(39)=1.5, p>.1). While we cannot determine whether this arises from the visual stimuli (picture vs. orthography), the lexical status of the auditory stimulus, or the presence of multiple competitors, it does suggest that the use of a 2AFC task alone is not sufficient to generate the artificially steep slope seen in Experiment 2.

Fixation Proportions—Distance from each participant's category boundary (rVOT) was used as the independent variable and binned in 5 ms increments. Although there was some variability in the category boundaries between participants, all 22 participants had data for rVOT values of −5, −10 and −15 ms, and 21 participants had data for an rVOT of −20 ms. We opted to analyze all four steps on the voiced side by removing one participant from analyses on this side of the continuum; all results were similar when data from that participant were included. On the voiceless side of the continuum, more rVOT values were available, with all 22 participants having data from +5 to +25 ms.

Figure 9 shows the effect of rVOT on competitor fixations. Separate one-way ANOVAs revealed a significant main effect of rVOT on both sides of the continuum when all of the tokens were included (**B:** $F_1(3,60)=13.6$, $\eta_P^2=.40$, $p<0.001$; $F_2(3,15)=28.7$, $\eta_P^2=0.85$, $p=0.0001$; **P:** $F_1(4,84)=8.7$, $\eta_P^2=.29$, $p=0.001$; $F_2(4,20)=11.8$, $\eta_P^2=0.7$, $p=0.0001$). Linear trends were also significant (**B:** $F_1(1,20)=29.1$, $\eta_P^2=.59$, $p<0.001$; $F_2(1,5)=43.9$, $\eta_P^2=0.9$, $p=0.001$; **P:** $F_1(1,21)=14.0$, $\eta_P^2=.40$, $p=0.001$; $F_2(1,5)=20.9$, $\eta_P^2=0.81$, $p=0.006$).

However, when tokens adjacent to the category boundary were removed from the analysis, sensitivity to within-category detail disappeared on the voiced, but not the voiceless, side of the continuum. There was no main effect or linear trend on the voiced side of the continuum (**B:** $F_1(2,40)=1.4$, $\eta_P^2=.07$, $p>0.1$; $F_2(2,10)=1.0$, $\eta_P^2=0.17$, $p>0.2$; Trend: $F_1(1,20)=2.3$, $\eta_P^2=.11$, $p>0.1$; $F_2(1,5)=0.5$, $\eta_P^2=0.09$, $p>0.2$). However, there was a significant main effect on the voiceless side (**P:** $F_1(3,63)=3.8$, $\eta_P^2=.15$, $p=0.014$; $F_2(3,15)=7.8$, $\eta_P^2=0.61$, $p=0.002$). The linear trend was not reliable by participants (**P:** $F_1(1,21)=2.2$, $\eta_P^2=.10$, $p>0.1$, but it was reliable by items (**P:** $F_2(1,5)=14.3$, $\eta_P^2=0.74$, $p=0.013$).

Fixation Durations—As before, heirarchical regressions were used to assess the effect of rVOT on the duration of the first look to the competitor. Participant codes were added in the first step to account for around 10% of the variance (**B:** $R^2=.108$, $F_{change}(21,953)=5.5$, $p=0.0001$; **P:** $R^2=.095$ $F_{change}(21,1345)=6.7$, $p=0.0001$). rVOT was added in the second step and while it only accounted for an additional .7% of the variance it was significant on the voiced side (**B:** $F_{change}(1,952)=7.2$, $p=0.007$). On the voiceless side, it accounted for 2.4% of the variance and was also significant (**P:** $F_{change}(1,1344)=36.5$, $p=0.0001$).

Results were similar when tokens less than 5 ms from the category boundary were excluded: participant codes accounted for an initial 10–15% of the variance (**B:** $R^2=.142$, $F_{change}(21,699)=5.5$, $p=0.0001$; **P:** $R^2=.097$, $F_{change}(21,1059)=5.4$, $p=0.0001$), and rVOT accounted for a small, but significant, amount of additional variance (**B:** $R^2_{change}=.006$, $F_{change}(1,698)=4.7$, $p=0.03$; **P:** $R^2=.011$ $F_{change}(1,1058)=12.5$, $p=0.0001$).

Summary—Evidence for gradiency on both the voiced and voiceless sides of the continuum was seen in the analyses of the fixation proportions when all of the tokens were included. However, when tokens adjacent to the category boundary were excluded, gradiency only remained on the voiceless side of the continuum. In addition, the duration analyses for first fixations found some evidence for gradiency on both sides of the continuum. Thus, as in the previous experiments, sensitivity to within-category detail can be seen but does not appear to be as robust in this 2AFC lexical task.

Discussion: Experiments 2–5

The goal of Experiments 2–5 was to understand why fixation patterns in picture identification tasks with words (as reported for synthesized speech stimuli in McMurray et al., 2002 and for modified natural speech stimuli in Experiment 1) show strong within-category gradient effects for VOT compared to the absence of gradiency in classic 2AFC phoneme identification tasks. One possibility was that one or more task or stimulus variables might account for the differences. A second possibility was that, in all speech tasks, eye movements might be a finer-grained measure of on-going decisions than button-press identification measures, revealing gradient effects that can be masked by explicit, discrete decision processes.

Experiment 2 used a two-alternative forced choice phoneme identification task with non-lexical CV stimuli. This is the type of task that has provided the strongest support for categorical perception and the absence of within-category sensitivity. The identification functions replicated the steep slopes indicative of categorical perception, and these slopes were much steeper than identification functions obtained from comparable lexical identification tasks or 4AFC phoneme decision tasks. These identification functions became less steep when four alternative phoneme decisions were obtained with CV stimuli (Experiment 3) or lexical stimuli (Experiment 4) and when two alternatives were used with words in a picture identification task (Experiment 5). This pattern of results suggests that both the number of response alternatives and the lexicality of the stimuli (or the task in the case of lexical stimuli whose initial phoneme was the focus of attention as in Experiment 4) affect the steepness of the identification function.

As we discussed earlier, it seems unlikely that differences in identification slope arise from the perception of VOT *prior to categorization*. Therefore we conclude that the 4AFC task is more sensitive to the continuous underlying representation of VOT allowing this to be seen in the labeling data (as a shallower slope).

In contrast to standard discrimination tasks, we assessed discrimination indirectly, by examining fixations to competitors (either pictures corresponding to CVs in Experiments 2 and 3 or words in Experiments 4 and 5). Within-category discrimination predicts gradient effects of rVOT on fixations to the pictured displays of the response competitors. Table 2 summarizes the effects of rVOT and the linear trends for Experiments 2–5 when tokens adjacent to the category boundary were included. Each experiment shows significant effects of rVOT and significant linear trends on both the voiced and voiceless ends of the continuum, with the exception of Experiment 2, where the effects of rVOT and the linear trend were not reliable for voiced tokens. Recall that the eye movement data come only from those trials where the response was consistent with the participant's category boundary—this is a conservative analysis. Thus these data support within-category gradient effects under most of the testing conditions.

The evidence for gradient effects is less definitive when we exclude tokens adjacent to the category boundary to provide the most stringent test of gradiency when the most ambiguous tokens are eliminated. Table 3 summarizes the effects of rVOT and the linear trend for each

experiment for these tokens. On the voiced side, the main effect of rVOT and the linear trend analyses did not reach significance for any of the four experiments, although the pattern was generally consistent with gradient sensitivity. On the voiceless side of the continuum, both the effects of rVOT and the linear trend were significant for phoneme identification of CV stimuli in the 2AFC task (Experiment 2) and for word stimuli in the 4AFC task (Experiment 4). The effects of both rVOT and the linear trend were not reliable for the 4AFC task with CV stimuli (Experiment 3) and for the 2AFC task with words (Experiment 5). Thus, while there are few clear differences that emerge between these tasks, the overall picture is that non-lexical tasks or tasks with fewer alternatives reveal gradient effects, but these effects are diminished compared to lexical tasks with more alternatives.

The overall pattern of more robust gradiency on the voiceless side can be explained in two ways. First, the analyses on the voiced side generally had less power than analyses on the voiceless side because the category boundaries fell between steps 4 and 5 along the 9-step continua. Thus, there were typically five rVOTs on the voiceless side of the continuum and four for nearly all participants after the token adjacent to the category boundary was removed. In contrast, the voiced side typically had only four rVOTs, leaving three after adjacent tokens were removed. Moreover, for Experiment 3, only three rVOTs could be analyzed, two after the adjacent token was removed.

Second, as measured in production studies, the distribution of VOTs in English is wider for voiceless than voiced sounds and the voiceless category tends to show larger changes due to phonetic context (Allen & Miller, 1999; Lisker & Abramson, 1964). Given the view that continuous perceptual inputs and stored categories interact in real-time during perception, it is quite reasonable that categories with different properties may yield different outcomes with respect to the maintenance of continuous detail. Thus, it is perhaps not surprising that the voiceless side of the continuum showed more gradient sensitivity to within-category detail than the voiced side.

Experiments 2–5 varied the nature of the stimulus (word or non-lexical CV), the number of alternatives (two or four) and the type of task (phoneme identification or picture identification). The combinations of variables represented in these four experiments do not explore the full parameter space of ($2 \times 2 \times 2 = 8$) possible combinations. Nonetheless, the pattern of results suggests that no single variable determines whether or not gradient effects are observed. However, each of the experiments showed trends that were partially supportive of gradient effects. This pattern of results is most consistent with the hypothesis that the underlying processing is gradient, with the effects emerging most clearly in lexical tasks with multiple (greater than 2) alternatives. Combinations of stimulus/task parameters that reduce the number of response alternatives, uses stimuli that are not words, or require a meta-linguistic rather than a lexical judgment seem to make it more difficult to observe gradient effects.

We also conducted a combined analysis across Experiments 2–5. If the underlying pattern is truly gradient, then the effect of rVOT on competitor activation and the linear trends should be more evident as statistical power is increased. In addition, the effect of VOT should not interact with experiment.

Before evaluating these predictions, it is important to note that because of the multiple differences between experiments, the presence or absence of an interaction must be interpreted with caution. The predicted non-significant interaction does not indicate that these four tasks are equivalent, (even less so than usual because of the multiple differences between experiments). Conversely, a significant interaction will not allow us to make claims about any particular task manipulations (since these five experiments vary on several

dimensions), but it would reveal that there are some robust differences in the gradient effects across the experiments. Thus, while our primary prediction is the overall main effect, the interaction term may at least be an indicator of any large task differences.

Our prediction of a main effect of VOT with no interaction with task was confirmed. In the combined analysis, there were significant effects of rVOT (**B:** $F(3,267)=42.3$, $\eta_P^2=.32$, $p<.001$; **P:** $F(4,396)=44.6$, $\eta_P^2=.31$, $p<.001$) and significant linear trends (**B:** $F(1,89)=76.1$, $\eta_P^2=.46$, $p<.001$; **P:** $F(1,99)=72.3$, $\eta_P^2=.42$, $p<.001$) when all tokens were included. The interaction with experiment was non-significant for voiceless sounds (**P:** $F<1$), and only marginally significant for voiced sounds (**B:** $F(12,267)=1.7$, $\eta_P^2=.07$, $p=.063$).

Most importantly, when the tokens adjacent to the category boundary were eliminated, reliable effects of rVOT were still present, both on the voiced side of the continuum (**B:** $F(2,178)=6.5$,, $\eta_P^2=.07$, $p=.002$; Linear Trend: $F(1,89)=9.9$, $\eta_P^2=.10$, $p=.002$) and on the voiceless side (**P:** $F(3,297)=16.1$, $\eta_P^2=.14$, $p<.001$; Linear Trend: $F(1,99)=19.6$, $\eta_P^2=.17$, $p<.001$), and there was no interaction of rVOT with experiment on either the voiced (**B:** $F(8,178)=1.7$, $\eta_P^2=.07$, $p=.09$) or voiceless sides (**P:** $F(12,297)=1.3$, $\eta_P^2=.05$, $p>.1$).

Thus, significant effects of rVOT and significant linear trends emerged in the combined analysis for both the P- and B-sides of the VOT continua, both when all tokens were included and when tokens adjacent to the category boundary were removed. None of the interactions of rVOT with experiment were significant suggesting that any differences were smaller than could be detected with the present analyses. Thus, although for any one of these experiments evidence for gradient sensitivity to differences in VOT for within-category stimuli, when tokens adjacent to the category boundary were excluded, is only partially reliable, the combined analysis is most consistent with gradient sensitivity. Again, it is important to note that we employed a very conservative test of gradiency by eliminating decisions that were inconsistent with the labeling data, and the strongest evidence of gradiency was obtained under conditions (lexical stimuli, lexical judgments, 4-alternatives) most similar to on-line processing in natural settings (lexical stimuli, lexical judgments, thousands of response alternatives).

Summary and Conclusions

The results presented here demonstrate that speech processing is sensitive to within-category acoustic variation along a VOT continuum. Experiment 1 found that the gradient effects reported with synthetic speech (McMurray et al., 2002) are also present with stimuli constructed from natural speech. This is an important finding because results with synthesized speech do not always hold for more natural stimuli. Fixation patterns to pictures of labeled objects revealed sensitivity to sub-phonetic detail even with an extremely conservative data-analytic approach that (a) filtered out “incorrect” trials; (b) removed between-participant category-boundary variability from the data prior to analysis; and (c) excluded tokens adjacent to the category boundary. Moreover, we have established that these effects can be seen using a continuous dependent measure (duration of fixation to the competitor). The finding that gradient effects occur with natural stimuli is important because results with synthesized speech do not always hold for more natural stimuli.

Experiments 2 – 5 manipulated task, number of response alternatives, and stimulus type to determine the generality of gradient sensitivity to within-category phonetic detail. These experiments provide a bridge between lexical tasks (e.g., the Visual World paradigm and cross-modal priming) that have shown sensitivity to within-category detail, and phoneme decision and discrimination tasks that have often found evidence for more categorical effects. It is clear that the evidence for gradient effects in Experiments 2 – 5 was weaker

than in experiments with words and four alternatives (McMurray et al., 2002; Experiment 1). However, each experiment showed some evidence for gradient sensitivity, including Experiment 2, which used a standard two-alternative forced-choice task and CV stimuli whose identification function closely approximated the idealized step-functions for categorical perception. More generally, however, effects of gradient sensitivity were strongest with lexical stimuli, a lexical task, and more than two response alternatives.

We acknowledge that the data from Experiments 2 through 5 are not inconsistent with a view in which sensitivity to sub-phonetic detail is restricted to VOTs near the category boundary. However, the gradiency hypothesis may apply more broadly for several reasons. First, as we have discussed, most of these experiments found some evidence of gradient effects, even when the tokens closest to the category boundary were removed. Second, even when the gradient effects were not reliable, the trends were in the predicted direction. Finally, gradient effects are consistent with the literature on phonetic prototypes (Miller, 1997), work showing sensitivity to allophonic variants (McLennan, Luce & Charles-Luce, 2003; Sumner and Samuel, 2004; Connine, 2004), and the emerging literature indicating that word recognition is extremely sensitive to even non-phonemic acoustic detail, as predicted by exemplar models (e.g., Pisoni, 1993; Goldinger, 1998; Pisoni, 1997 for a review).

A final concern is that the strongest effects of gradiency are obtained in Visual World eye-tracking tasks with a small number of pictured referents. Thus, it could be argued that participants might adopt a task-specific strategy that takes advantage of the closed set of displayed items, perhaps by implicitly naming one of the pictured referents, thereby creating a verifications set that bypasses normal lexical processing. There is, of course no way to rule out such an explanation for any given experiment. However, it seems likely that simplifying the task in this way would result in more categorical responding. More importantly, there is now a body of research using the Visual World paradigm that is inconsistent with either implicit naming or a task-specific strategy that bypasses normal lexical processing. The most compelling evidence comes from studies demonstrating effects of cohort density and neighborhood density, even when these competitors are neither mentioned nor displayed (Magnuson et al., in press, for review and discussion, see Tanenhaus, in press).

If the goal of spoken language processing were to optimize mapping of the acoustic input onto phonetic and phonemic categories, then gradient sensitivity might not be desirable. However, natural language understanding operates in the context of decoding the acoustic cues that refer to one of thousands of lexical alternatives. Lexical access must operate in real time at a rate that places extreme demands on processing efficiency to keep pace with the nearly continuous flow of fluent speech. Thus, to maximize efficiency under such conditions, it is imperative to both make rapid temporary inferences about upcoming events, even if those predictions are based on partial information, and also to preserve partially ambiguous material that may be disambiguated by later information. The task of identifying words, therefore, may benefit from sensitivity to sub-phonetic detail, particularly if that sub-phonetic detail is predictive of upcoming information in the speech stream or can resolve ambiguity in previously heard material.

A number of studies have begun to show that this is the case: sensitivity of lexical activation to sub-phonetic detail is advantageous during on-line spoken word recognition. For example, place assimilation can create sub-phonetic modifications to coronal segments such that they are produced partially labialized (in the context of a labial stop consonant). Thus, the phrase *green boat* would be pronounced with a labialized coronal as *green*$_m$ *boat*. If activation for *boat* were sensitive to this within-category modification, the system would be able to anticipate its occurrence on the basis of this preceding input. Evidence for these progressive effects has been reported in a number of recent studies (Gow, 2001; Gow, 2002; Gow &

McMurray, in press). Moreover, this relationship between systematic within-category variation and lexical activation has now been documented in a number of other domains, including sensitivity to vowel length in segmenting embedded words (Gow & Gordon, 1995; Salverda, et al., 2003, 2007), dealing with mispronunciation (Goldrick & Blumstein, 2006; McMurray, 2004), short-term integration of phonetic cues (McMurray, Clayards, Aslin & Tanenhaus, 2004), and revision of prior commitments to recover from lexical garden-paths (McMurray, Tanenhaus & Aslin, submitted). All of these results suggest that sensitivity to fine-grained acoustic detail may enhance word recognition by reducing ambiguity, anticipating upcoming events, and aiding in the integration of acoustic/phonetic material over time.

Acknowledgments

The authors would like to thank Katherine Crosswhite for the amplitude normalization script. This work was supported by NIH grants F31DC006537-01 to BM, and DC005071 to MKT and RNA.

References

Allen JS, Miller JL. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. Journal of the Acoustical Society of America 1999;106:2031–2039. [PubMed: 10530026]

Allen JS, Miller JL. Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. Perception & Psychophysics 2001;63(5):798–810. [PubMed: 11521848]

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye-movements: evidence for continuous mapping models. Journal of Memory and Language 1998;38(4):419–439.

Anderson JA, Silverstein JW, Ritz SA, Jones RS. Distinctive features, categorical perception and probability learning: some applications of a neural model. Psychological Review 1977;84:413–451.

Andruski JE, Blumstein SE, Burton MW. The effect of subphonetic differences on lexical access. Cognition 1994;52:163–187. [PubMed: 7956004]

Beale JM, Keil FC. Categorical effects in the perception of faces. Cognition 1995;57(3):217–239. [PubMed: 8556842]

Bornstein MH, Korda NO. Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. Psychological Research 1984;46(3):207–222. [PubMed: 6494375]

Burton MW, Blumstein SE. Lexical effects on phonetic categorization: the role of stimulus naturalness and stimulus quality. Journal of Experimental Psychology: Human Perception and Performance 1995;21(5):1230–1235. [PubMed: 7595247]

Carney AE, Widin GP, Viemeister NF. Non categorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America 1977;62:961–970. [PubMed: 908791]

Connine CM. 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 and Review 2004;11(6): 1084–1089. [PubMed: 15875980]

Cooper R. The control of eye fixation by the meaning of spoken language. Cognitive Psychology 1974;6:84–107.

Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology 2001a;42:317–367. [PubMed: 11368527]

Dahan D, Magnuson JS, Tanenhaus MK, Hogan E. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes 2001b; 16(5/6):507–534.

Dahan D, Tanenhaus MK. Continuous mapping from sound to meaning in spoken language comprehension: Evidence from immediate effects of verb-based constraints. Journal of Experimental Psychology: Learning, Memory and Cognition 2004;30:498–513.

Damper RI, Harnad SR. Neural network models of categorical perception. Perception & Psychophysics 2000;62(4):843–867. [PubMed: 10883589]

Davis M, Marslen-Wilson W. Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance 2002;28(1):218–244.

Ferrero FE, Pelamatti GM, Vagges K. Continuous and categorical perception of a fricative-affricate continuum. Journal of Phonetics 1982;10(3):231–244.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representations of visual stimuli in the primate prefrontal cortex. Science 2001;291(5502):312–316. [PubMed: 11209083]

Fry DB, Abramson AS, Eimas PD, Liberman AM. The identification and discrimination of synthetic vowels. Language and Speech 1962;5:171–189.

Ganong WF. Phonetic categorization in auditory word recognition. Journal of Experimental Psychology: Human Perception and Performance 1980;6(1):110–125. [PubMed: 6444985]

Gerrits E, Schouten MEH. Categorical perception depends on the discrimination task. Perception & Psychophysics 2004;66(3):363–376. [PubMed: 15283062]

Goldinger SD. Echoes of echos? An episodic theory of lexical access. Psychological Review 1998;105(2):251–279. [PubMed: 9577239]

Goldrick M, Blumstein. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. Language and Cognitive Processes 2006;21:649–683.

Goldstone RL, Lippa Y, Shiffrin RM. Altering object representations through category learning. Cognition 2001;78:27–43. [PubMed: 11062321]

Gow D. Assimilation and anticipation in continuous spoken word recognition. Journal of Memory and Language 2001;45:133–139.

Gow D. Does English coronal place assimilation create lexical ambiguity? Journal of Experimental Psychology: Human Perception and Performance 2002;28(1):163–179.

Gow D, Gordon P. Lexical and prelexical influences on word segmentation: Evidence from priming. Journal of Experimental Psychology: Human Perception & Performance 1995;21(2):344–359. [PubMed: 7714476]

Gow D, McMurray B. Word recognition and phonology: The case of English coronal place assimilation. Papers in Laboratory Phonology :IX. (in press).

Healy AF, Repp B. Context independence and phonetic mediation in categorical perception. Journal of Experimental Psychology: Human Perception and Performance 1982;8(1):68–80. [PubMed: 6460086]

Howard DM, Rosen S, Broad V. Major/minor triad identification and discrimination by musically trained and untrained listeners. Music Perception 1992;10(2):205–220.

Klatt D. Software for a cascade/parallel synthesizer. Journal of the Acoustical Society of America 1980;67:971–995.

Kopp J. A new test for categorical perception. Psychological Record 1969;19(4):573–578.

Kuhl PK. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. Perception & Psychophysics 1991;50:93–107. [PubMed: 1945741]

Ladefoged, P. A course in phonetics. New York: Harcourt Brace Publishers; 1993.

Larkey L, Wald J, Strange W. Perception of synthetic nasal consonants in initial and final syllable position. Perception & Psychophysics 1978;23(4):299–312. [PubMed: 748852]

Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 1957;54(5):358–368. [PubMed: 13481283]

Liberman AM, Harris KS, Kinney J, Lane H. The discrimination of relative onset-time of the components of certain speech and non-speech patterns. Journal of Experimental Psychology 1961;61:379–388. [PubMed: 13761868]

Lindblom B. Role of articulation in speech perception: Clues from production. Journal of the Acoustical Society of America 1996;99(3):1683–1692. [PubMed: 8819859]

Lisker L, Abramson AS. A cross-language study of voicing in initial stops: acoustical measurements. Word 1964;20:384–422.

Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. Cognitive Science 2007;31:133–156.

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The microstructure of spoken word recognition: Studies with artificial lexicons. Journal of Experimental Psychology: General 2003;133(2):202–227. [PubMed: 12825637]

Marslen-Wilson W, Warren P. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. Psychological Review 1994;101(4):653–675. [PubMed: 7984710]

Massaro DW, Cohen MM. Categorical or continuous speech perception: a new test. Speech Communication 1983;2:15–35.

McClelland J, Elman J. The TRACE model of speech perception. Cognitive Psychology 1986;18(1):1–86. [PubMed: 3753912]

McClelland JL, Fiez JA, McCandliss BD. Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. Physiology & Behavior 2002;77:657–662. [PubMed: 12527015]

McLennan C, Luce PA, Charles-Luce J. Representation of lexical form. Journal of Experimental Psychology: Learning, Memory and Cognition 2003;29(4):539–553.

McMurray, B. Unpublished doctoral dissertation. Department of Brain and Cognitive Sciences, University of Rochester; 2004. Within-category variation is used in spoken word recognition: Temporal integration at two time scales.

McMurray. KlattWorks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research. (in preparation)

McMurray, B.; Clayards, M.; Aslin, RN.; Tanenhaus, MK. Gradient sensitivity to acoustic detail and temporal integration of phonetic cues. Poster presented at the 75th Meeting of the Acoustical Society of America; New York, NY. 2004 May.

McMurray B, Spivey MJ. The categorical perception of consonants: The Interaction of Learning and Processing. Proceedings of the Chicago Linguistics Society 1999;35:205–219.

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition 2002;86(2):B33–B42. [PubMed: 12435537]

McMurray, B.; Tanenhaus, MK.; Aslin, RN. Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. (submitted)

McMurray B, Tanenhaus MK, Aslin RN, Spivey M. Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. Journal of Psycholinguistic Research 2003;32(1):77–97. [PubMed: 12647564]

McQueen JM, Norris D, Cutler A. Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. Journal of Experimental Psychology: Human Perception and Performance 1999;25:1363–1389.

Miller JL. Internal structure of phonetic categories. Language and Cognitive Processes 1997;12:865–869.

Miller J, Connine C, Schermer T, Kluender K. A possible auditory basis for internal structure of phonetic categories. The Journal of the Acoustical Society of America 1983;73(6):2124–2133. [PubMed: 6875098]

Miller JL, Wayland SC. Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. Perception & Psychophysics 1993;54:205–210. [PubMed: 8361836]

Miller JL, Volaitis LE. Effect of speaking rate on the perceptual structure of a phonetic category. Perception & Psychophysics 1989;46(6):505–512. [PubMed: 2587179]

Newell FN, Bulthoff HH. Categorical perception of familiar objects. Cognition 2002;85:113–143. [PubMed: 12127696]

Philips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel DM, McGinnis Roberts T. Auditory cortex accesses phonological categories: An MEG Mismatch Study. Journal of Cognitive Neuroscience 2000;12(6):1038–1055. [PubMed: 11177423]

Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics 1973;13(2):253–260.

Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. Speech Communication 1993;13:109–125.

Pisoni, DB. Some thoughts on “normalization” in speech perception. In: Johnson, K.; Mullenix, JW., editors. Talker variability in speech processes. San Diego, CA: Academic Press; 1997. p. 9-32.

Pisoni DB, Aslin RN, Perey AJ, Hennessy BL. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. Journal of Experimental Psychology: Human Perception and Performance 1982;8(2):297–314. [PubMed: 6461723]

Pisoni DB, Lazarus JH. Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America 1974;55(2):328–333. [PubMed: 4821837]

Pisoni DB, Tash J. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics 1974;15(2):285–290.

Quinn P. Visual perception of orientation is categorical near vertical and continuous near horizontal. Perception 2004;33(8):897–906. [PubMed: 15521689]

Repp, BH. Categorical perception: Issues, methods, findings. In: Lass, NJ., editor. Speech and language: Advances in basic research and practice. Academic Press; San Diego, CA: 1984. p. 243-335.

Salverda AP, Dahan D, McQueen J. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition 2003;90(1):51–89. [PubMed: 14597270]

Salverda AP, Dahan D, Tanenhaus MK, Crosswhite K, Masharov M, McDonough J. Effects of prosodically modulated sub-phonetic variation on lexical competition. Cognition 2007;105 (2): 466–476. [PubMed: 17141751]

Samuel A. The effect of discrimination training on speech perception: Noncategorical perception. Perception & Psychophysics 1977;22(4):312–330.

Samuel A. Phonetic prototypes. Perception & Psychophysics 1982;31(4):307–314. [PubMed: 7110883]

Schouten MEH, Van Hessen. Modeling phoneme perception I: Categorical Perception. Journal of the Acoustical Society of America 1992;92(4):1841–1855. [PubMed: 1401529]

Schouten B, Gerrits E, Van Hessen. The end of categorical perception as we know it. Speech Communication 2003;41(2003):71–80.

Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. Journal of the Acoustical Society of America 1999;106(2):1078–1083. [PubMed: 10462812]

Shinn P, Blumstein SE, Jongman A. Limits of context conditioned effects in the perception of [b] and [w]. Perception & Psychophysics 1985;38(5):397–407. [PubMed: 3831918]

Spivey, M. The continuity of thought. Oxford, UK: The Oxford University Press; 2006.

Streeter L, Nigro G. The role of medial consonant transitions in word perception. Journal of the Acoustical Society of America 1979;65:1533–1541. [PubMed: 489823]

Sumner M, Samuel A. Perception and representation of regular variation: The case of final /t/. Journal of Memory and Language 2005;52:322–338.

Tanenhaus, MK. Eye movements and spoken language processing. In: van Gompel, RPG.; Fischer, MH.; Murray, WS.; Hill, RL., editors. Eye movements: A window on mind and brain. Oxford: Elsevier; 2007. p. 4430-470.

Tanenhaus MK, Magnuson JS, Dahan D, Chambers CG. Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. Journal of Psycholinguistic Research 2000;29:557–580. [PubMed: 11196063]

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JE. Integration of visual and linguistic information in spoken language comprehension. Science 1995;268:1632–1634. [PubMed: 7777863]

Utman JA, Blumstein SE, Burton MW. Effects of subphonetic and syllable structure variation on word recognition. Perception & Psychophysics 2000;62(6):1297–1311. [PubMed: 11019625]

Van Hessen AJ, Schouten MEH. Categorical perception as a function of stimulus quality. Phonetica 1999;56:56–72. [PubMed: 10450076]

Warren P, Marslen-Wilson W. Continuous uptake of acoustic cues in spoken word recognition. Perception & Psychophysics 1988;41(3):262–275. [PubMed: 3575084]

Appendix A

KlattWorks scripts for stimuli used in McMurray et al (2002) and Experiment 2.

KlattWorks is freely available upon request of the author, and makes use of scripts to describe auditory stimuli. Each script records a series of linear and nonlinear operations that describe how specific Klatt parameters change over time. Please see www.psychology.uiowa.edu/faculty/mcmurray/klattworks for more information or McMurray (in preparation).

Provided here are complete scripts for each of the six voiced target words (beach, bear, bale, bomb, bump and butter). The final script shows an example of how the VOT continua was constructed from beach. The other five continua were constructed using identical scripts.

Beach

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	comment								Burst
3	logistic	19	29	AF	77.5	2.5	−2	33	
4	setto	19	29	AB	45				
5	setto	19	29	A2	30				
6	comment								Vowel
7	logistic	19	210	AV	60	55	−2	29	
8	logistic	51	210	AV	55	0	−20	61	
9	comment								F1
10	logistic	1	210	F1	280	180	80	20	
11	comment								F2
12	logistic	1	210	F2	2250	1550	80	20	
13	logistic	40	210	F2	2250	1750	−80	57	
14	comment								F3
15	logistic	1	210	F3	2890	2190	80	20	
16	comment								F0
17	logistic	1	66	F0	130	0	13	11	
18	logistic	26	118	F0	130	100	−1	54	
19	setto	1	210	FNZ	250				
20	comment								burst
21	logistic	70	92	AH	80	50	−1	70	
22	setto	70	90	AB	50				
23	logistic	70	90	AF	60	45	−2	80	
24	setto	70	90	A2	45				

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
25	setto	70	90	A3	45				
26	setto	70	90	A4	45				
27	setto	70	90	A5	15				
28	setto	70	90	A6	35				

Bale

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	comment								Burst
3	logistic	19	29	AF	77.5	2.5	−2	33	
4	setto	19	29	A2	30				
5	comment								Vowel
6	logistic	19	210	AV	60	55	−2	29	
7	logistic	60	210	AV	55	0	−20	86	
8	comment								F1
9	logistic	1	210	F1	480	180	80	20	
10	logistic	37	210	F1	480	440	−10	45	
11	comment								F2
12	logistic	1	210	F2	1600	900	80	20	
13	logistic	32	210	F2	1750	1600	15	40	
14	logistic	59	210	F2	1750	950	−100	69	
15	comment								F3
16	logistic	1	210	F3	2350	1650	80	20	
17	logistic	32	210	F3	2450	2350	20	40	
18	logistic	59	210	F3	3000	2450	90	69	
19	comment								F0
20	logistic	1	66	F0	130	0	13	11	
21	logistic	26	118	F0	130	100	−1	54	

Bear

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	logistic	19	29	AF	77.5	2.5	−2	33	
3	setto	19	29	AB	45				
4	setto	19	29	A2	30				
5	comment								Vowel
6	logistic	19	210	AV	60	55	−2	29	

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
7	logistic	55	210	AV	55	0	−20	78	
8	comment								F1
9	logistic	1	210	F1	480	180	80	20	
10	logistic	32	210	F1	480	440	−10	40	
11	comment								F2
12	logistic	1	210	F2	1600	900	80	20	
13	logistic	32	210	F2	1750	1600	15	40	
14	logistic	54	210	F2	1750	1310	−75	64	
15	comment								F3
16	logistic	1	210	F3	2350	1650	80	20	
17	logistic	32	210	F3	2450	2350	20	40	
18	logistic	54	210	F3	2450	1540	−90	64	
19	comment								F0
20	logistic	1	66	F0	130	0	13	11	
21	logistic	26	118	F0	130	100	−1	54	

Bomb

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	comment								Burst
3	logistic	19	29	AF	77.5	2.5	−2	33	
4	setto	19	29	AB	45				
5	setto	19	29	A2	30				
6	comment								Voicing
7	logistic	19	210	AV	60	55	−2	29	
8	logistic	66	210	AV	55	0	−20	79	
9	comment								F1
10	logistic	1	210	F1	710	300	80	20	
11	logistic	52	210	F1	710	510	−10	61	
12	comment								F2
13	logistic	1	210	F2	1000	500	80	20	
14	logistic	52	210	F2	1000	900	−10	61	
15	logistic	79	210	F2	1100	900	80	77	
16	comment								F3
17	logistic	1	210	F3	2540	1840	80	20	
18	comment								F0
19	logistic	1	66	F0	130	0	13	11	
20	logistic	26	210	F0	130	100	−1	54	

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
21	logistic	1	210	FNZ	500	250	40	61	

Bump

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	comment								burst
3	logistic	19	29	AF	77.5	2.5	−2	33	
4	setto	19	29	AB	45				
5	setto	19	29	A2	30				
6	comment								Voicing
7	logistic	19	210	AV	60	55	−2	29	
8	logistic	61	210	AV	55	0	−20	71	
9	comment								F1
10	logistic	1	210	F1	650	240	80	20	
11	logistic	33	210	F1	650	450	−10	51	
12	comment								F2
13	logistic	1	210	F2	1100	600	80	20	
14	logistic	33	210	F2	1100	1000	−10	51	
15	logistic	60	210	F2	1200	1000	80	67	
16	Comment								F3
17	setto	1	210	F3	2540				
18	Comment								F0
19	logistic	1	66	F0	130	0	13	11	
20	logistic	26	210	F0	130	100	−1	54	
21	logistic	1	210	FNZ	500	250	40	51	
22	logistic	71	210	FNZ	500	250	−40	82	
23	comment								burst: p
24	logistic	78	84	AH	97.5	22.5	−0.6	86	
25	setto	78	84	AB	50				
26	setto	78	84	AF	63				
27	setto	78	84	A2	30				

Butter

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
1	blank		210						
2	comment								burst
3	logistic	19	29	AF	77.5	2.5	−2	33	

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment
4	setto	19	29	AB	45				
5	setto	19	29	A2	30				
6	comment								voicing
8	logistic	19	210	AV	60	55	−2	29	
9	logistic	46	210	AV	55	0	−30	53	
10	logistic	54	210	AV	55	0	100	55	
11	logistic	69	210	AV	55	0	−20	89	
12	comment								F1
13	logistic	1	210	F1	650	240	80	20	
14	logistic	51	210	F1	650	470	−40	56	
15	comment								F2
16	logistic	1	210	F2	1100	600	80	20	
17	logistic	51	210	F2	1200	1100	50	56	
18	comment								F3
19	logistic	1		F3	2540	1840	80	20	
20	logistic	44	210	F3	2540	1540	−175	56	
21	comment								F0
22	logistic	1	66	F0	130	0	13	11	
23	logistic	26	118	F0	130	100	−1	54	
26	setto	1	210	FNZ	250				
27	comment								Burst (t)
28	setto	55	60	AH	65				
29	setto	51	55	AB	55				
30	setto	51	55	AF	60				

The final script shows an example of how the VOT continua was constructed from the target word *beach*. In each case, first the original parameters from *beach* were copied. Next a short segment of the initial voicing (AV) was set to 0. Finally, during that same time period, aspiration (AH) was set to 63 dB. To make the next step of the continuum, the *end* time of this period of time was systematically increased by 1 frame (5 ms.). No other parameters were changed.

All six continua were produced with identical scripts.

Beach/Peach Continua: 5 ms VOT

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment	From Word
1	blank		210							
2	copy									Beach
3	setto	19	19	AV	0					
4	setto	19	19	AH	63					

Beach/Peach Continua: 10 ms VOT

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment	From Word
1	blank		210							
2	copy									Beach
3	setto	19	20	AV	0					
4	setto	19	20	AH	63					

Beach/Peach Continua: 40 ms VOT

Order	Command	Start	End	Param.	Num1	Num2	Num3	Num4	Comment	From Word
1	blank		210							
2	copy									Beach
3	setto	19	26	AV	0					
4	setto	19	26	AH	63					

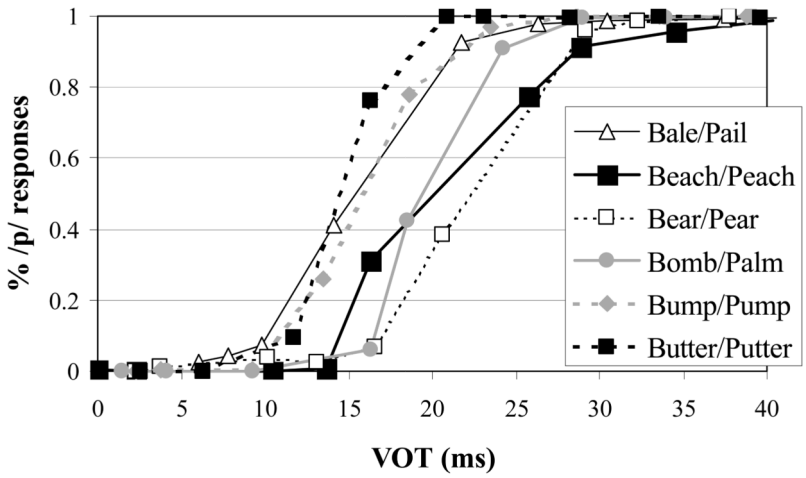

Figure 1.
Identification functions obtained from mouse responses in Experiment 1. Shown are the proportion of tokens labeled as /p/ as a function of VOT. Note the variability in category boundaries between continua.

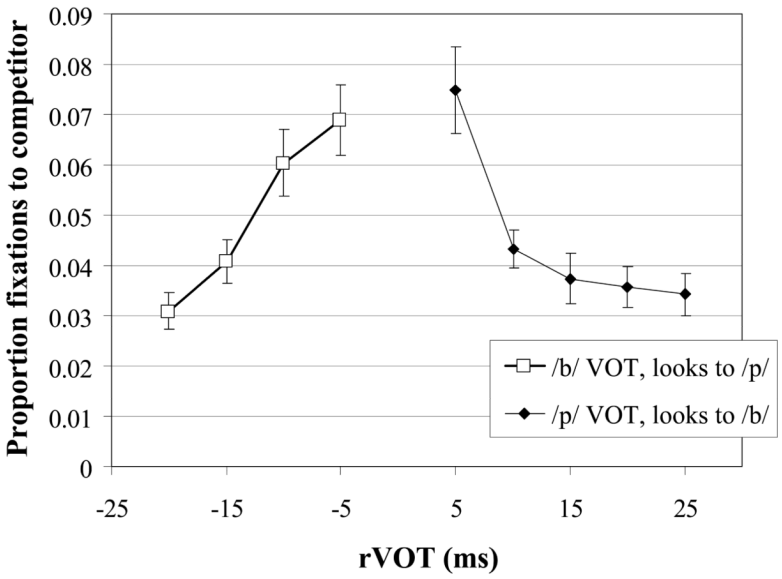

Figure 2.
Proportion of fixations to the competitor as a function of VOT for voiced (left series) and voiceless (right series) tokens. rVOT's were grouped into 5-ms bins (equivalent to the grouping performed in the second ANOVA). Error bars reflect SEM.

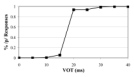

Figure 3.
Percentage of trials in which /p/ was selected in Experiment 2 and in McMurray et al (2002). A clear difference in the slope of the function can be seen.

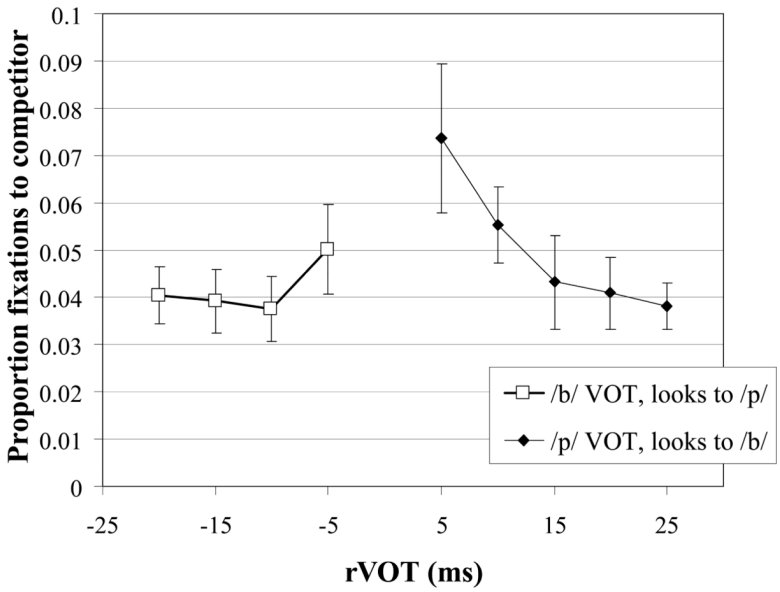

Figure 4.
Proportion of fixations to the competitor as a function of rVOT for voiced (left series) and voiceless (right series) tokens in Experiment 2. Error bars reflect SEM.

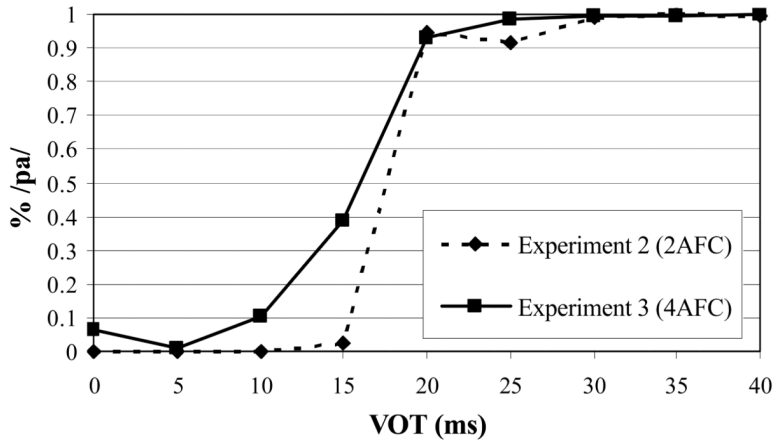

Figure 5.
Percentage of trials in which /p/ was selected as a function of VOT. Experiment 3 is contrasted with Experiment 2. The category boundaries are very similar, but the slope for Experiment 3 was quite a bit steeper than Experiment 2.

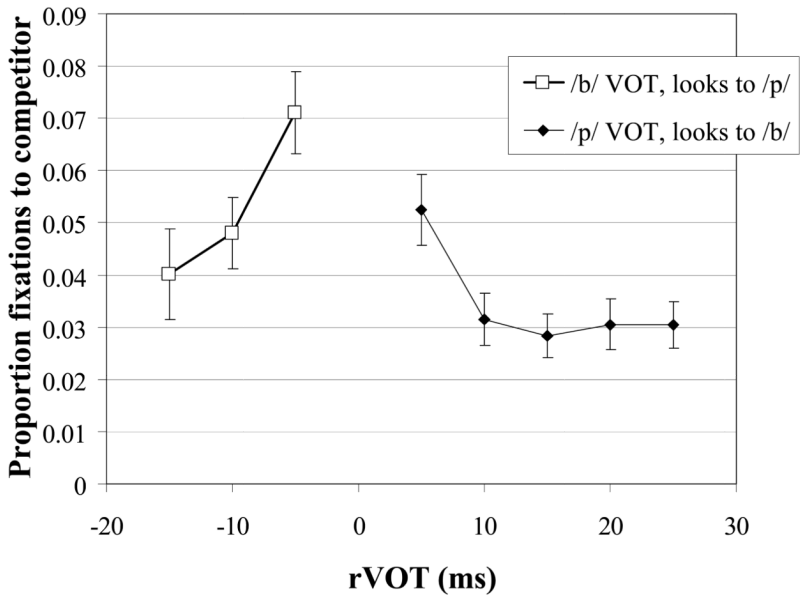

Figure 6.
Proportion of fixations to the competitor as a function of rVOT in Experiment 3. Error bars reflect SEM.

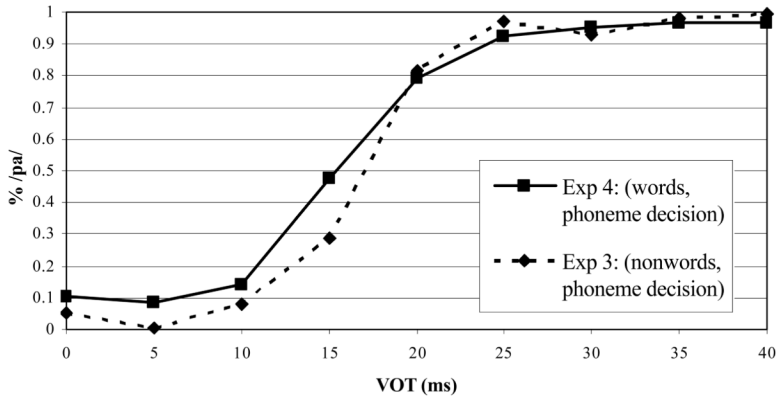

Figure 7.

Percentage of trials in which /p/ was selected as a function of VOT and experiment for three 4AFC tasks. Experiment 4 (4AFC phoneme decision on lexical stimuli) is contrasted with Experiment 3 (4AFC phoneme decision on CV stimuli).

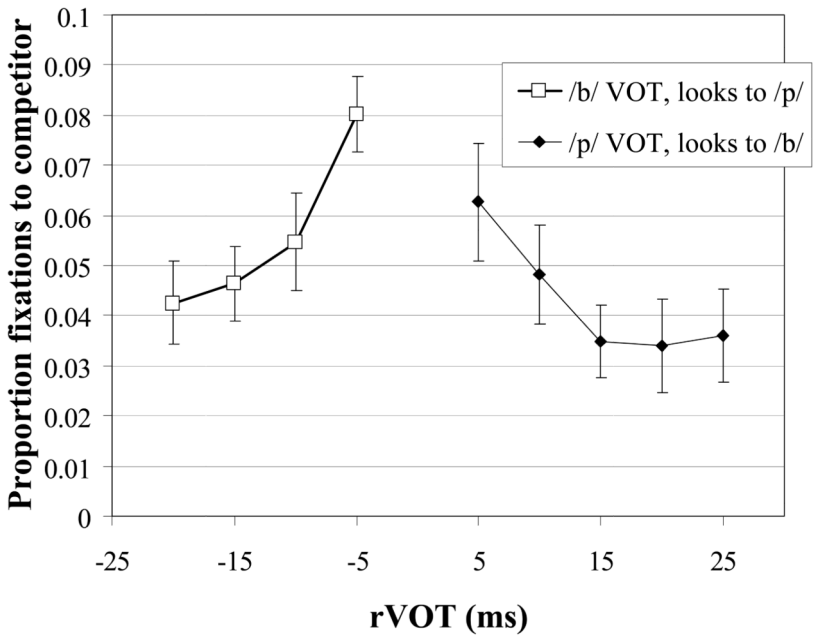

Figure 8.
Probability of fixating the competitor as a function of rVOT in Experiment 4. Error bars reflect SEM.

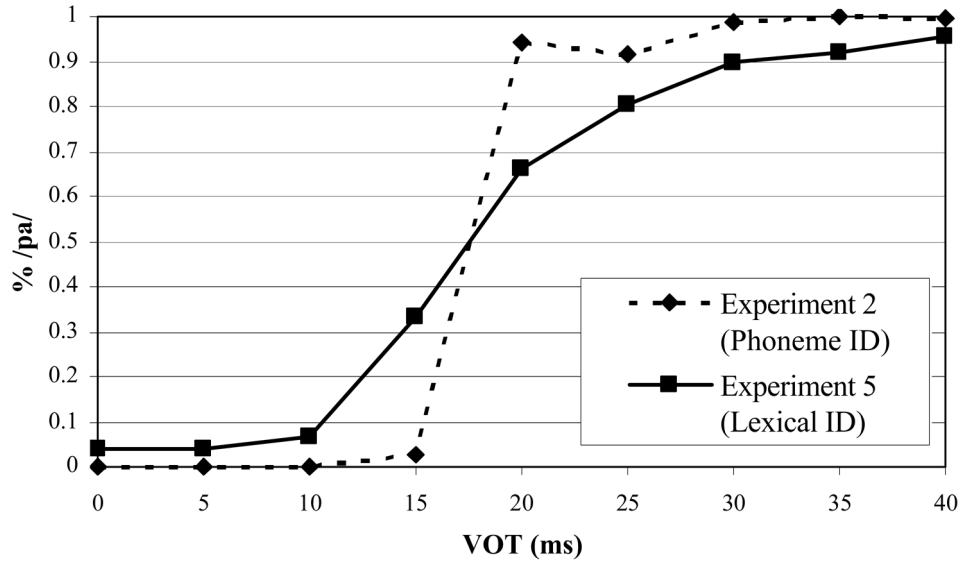

Figure 9.
Proportion of trials identified as /p/ as a function of VOT for Experiment 2 (2AFC phoneme ID) and Experiment 5 (2AFC lexical ID).

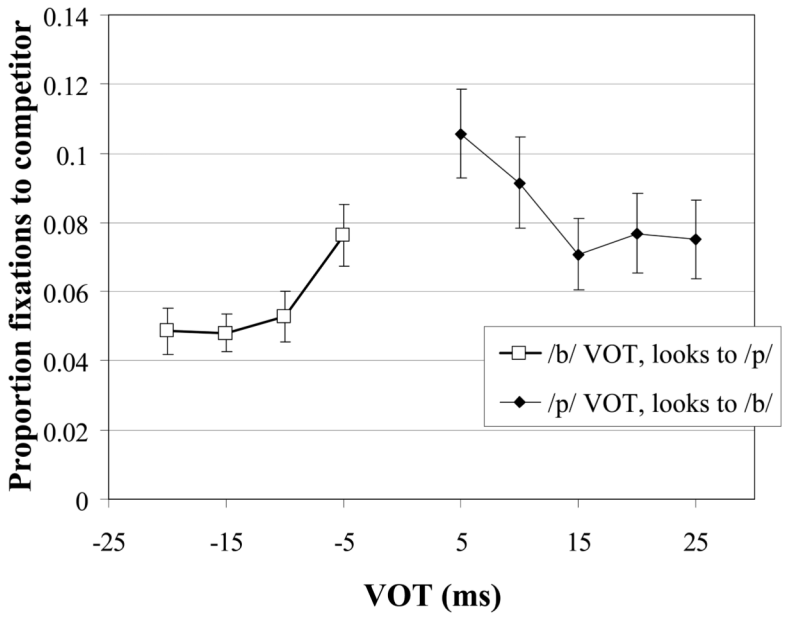

Figure 10.
Probability of fixating the competitor as a function of rVOT in Experiment 5. Error bars reflect SEM.

Table 1

VOTs of the cross-spliced stimuli used in Experiment 1. Items are categorized by overall tendency (across participants) to categorize as /b/ or /p/ based on mouse-click responses. This category boundary does not necessarily correspond to the boundary used during analysis (which was computed within-participant and item).

	Bale	**Beach**	**Bear**	**Bomb**	**Bump**	**Butter**
Voiced VOTs(ms)			0			
		0	4.6	4	2.3	
	0	4.8	10.4	9	4.8	
	5.7	12.2	14.4	9.7	11.4	0
	15.1	14.9	17.6	15.7	14.5	7.9
	15.7	20.5	22.4	17.5	18.9	11.4
Observed Category Boundary						
Voiceless VOTs (ms)	23.3	27.7	30.2	23.6	25	18.4
	27.7	29.9	32.8	29.7	29	24.3
	30.8	35.6	39.4	32.6	33.9	25.4
	41.6	41.7		38.7	43.3	33.8
	42.9					35.5
						41.3

J Exp Psychol Hum Percept Perform. Author manuscript; available in PMC 2010 December 29.

Table 2

Summary of results from Experiment 1–5 when all tokens were included in the analyses. Given are η_P^2's (effect sizes) resulting from one-way analyses of variance using rVOT as an independent factor and the proportion of looks to the competitor as a dependent measure. Complete descriptions of these analyses are provided in the results section for each experiment.

Experiment	Voiced Side		Voiceless	
	M.E.	Trend	M.E.	Trend
1 (4AFC words)	.54**	.69**	.48**	.60**
2 (2AFC CV)	.13+	.14+	.25**	.36**
3 (4AFC CV)	.28**	.39**	.24**	.30**
4 (4AFC words)	.38**	.57**	.40**	.49**
5 (2AFC words)	.40**	.59**	.29**	.40**

+ p<.10;

* p<.05;

** p<.01

Table 3

Summary of results from Experiment 1–5 when tokens adjacent to the category boundary were removed from the analysis. Given are η_p^2's (effect size) resulting from one-way analyses of variance using rVOT as an independent factor and the proportion of looks to the competitor as a dependent measure. Complete descriptions of these analyses are provided in the results section for each experiment.

Experiment	Voiced Side		Voiceless	
	M.E.	Trend	M.E.	Trend
1 (4AFC words)	.48**	.63**	.097	.336**
2 (2AFC CV)	.02	.03	.15*	.33**
3 (4AFC CV)	.07	.07	.01	.00
4 (4AFC words)	.08	.12	.23**	.24*
5 (2AFC words)	.06	.10	.15*	.10

* p<.05;

** p<.01.