

Tracking the time course of phonetic cue integration during spoken word recognition

BOB McMURRAY

University of Iowa, Iowa City, Iowa

MEGHAN A. CLAYARDS

University of York, York, England

AND

MICHAEL K. TANENHAUS AND RICHARD N. ASLIN

University of Rochester, Rochester, New York

Speech perception requires listeners to integrate multiple cues that each contribute to judgments about a phonetic category. Classic studies of trading relations assessed the weights attached to each cue but did not explore the time course of cue integration. Here, we provide the first direct evidence that asynchronous cues to voicing (/b/ vs. /p/) and manner (/b/ vs. /w/) contrasts become available to the listener at different times during spoken word recognition. Using the visual world paradigm, we show that the probability of eye movements to pictures of target and of competitor objects diverge at different points in time after the onset of the target word. These points of divergence correspond to the availability of early (voice onset time or formant transition slope) and late (vowel length) cues to voicing and manner contrasts. These results support a model of cue integration in which phonetic cues are used for lexical access as soon as they are available.

Spoken word recognition requires that the perceptual system cope with a noisy signal, sequential inputs that persist for only a fleeting moment, and temporary ambiguity as words unfold. Two particularly challenging aspects of this requirement are that phonemic and lexical contrasts are rarely instantiated along a single dimension (cue) and information from disparate cues often arrives asynchronously. Therefore, at any point in time, the word recognition system rarely has all of the relevant information for a phonetic distinction.

Integration of asynchronous cues has not been a focus of prior theories of spoken word recognition. One approach would be for the system to store early cues in a temporary buffer until the remaining cues have arrived. This would minimize premature (incorrect) commitments but would force the system to delay making even a partial decision. Alternatively, each cue could provide partial evidence for higher level units (e.g., features, phonemes, or words) as soon as it arrived. This would speed up recognition if early commitments are correct but risks delay if they must be revised. Of course, the system could buffer some cues but treat others continuously; buffers may exist at multiple levels of representation, and some buffers might release preliminary analyses for higher level processing.

We examined two phonetic distinctions that are determined by asynchronous cues in order to find out whether lexical activation is immediately sensitive to the early

information or is delayed until both cues are available. Existing models have not addressed this question directly. The fuzzy logical model of perception (Oden & Massaro, 1978) assumes that cues are simultaneously available (which would be true in a buffered system), but the model could likely function as an immediate integrator by treating cues that have not arrived yet as ambiguous (Oden, personal communication, May 2008). TRACE (McClelland & Elman, 1986), which is clearly consistent with continuous integration, makes the simplifying assumption that cues are available simultaneously, but it is unclear whether this is necessary for the dynamics of the model. There has also been debate about asynchrony in models of feature parsing. Gow (2003) argued that cues must be buffered and integrated, whereas Fowler (1984) argued for continuous integration. Thus, answering this question would have implications for a number of approaches to cue integration.

We examined two word-initial consonant contrasts (voicing and manner) that are first cued by information at word onset and later by the length of the vowel. Word-initial voicing (e.g., /p/ vs. /b/) is cued by voice onset time (VOT), along with other cues (e.g., pitch and first formant frequency). Vowel length varies systematically with voicing (Allen & Miller, 1999; Kessinger & Blumstein, 1998), and it participates in a trading relation with VOT (J. L. Miller & Volaitis, 1989; Summerfield, 1981); an

B. McMurray, bob-mcmurray@uiowa.edu

ambiguous VOT is perceived as a voiced consonant when followed with a long vowel and as a voiceless consonant with a short vowel.

Word-initial manner of articulation (e.g., /b/ vs. /w/) is cued by the slope of the formant transitions. For medium transitions, the likelihood of perceiving a /w/ increases with short vowels (J. L. Miller & Liberman, 1979; J. L. Miller & Wayland, 1993; but see Shinn, Blumstein, & Jongman, 1985). Effects of vowel length can be separated from effects of sentential speaking rate (Summerfield, 1981; Wayland, Miller, & Volaitis, 1994) and can be observed across a range of rates (Allen & Miller, 1999). Thus, vowel length appears to operate as an independent cue (Repp, 1982).

Although numerous studies have examined the final product of cue integration (see Repp, 1982, for a classic review), few have assessed the time course of integrating asynchronous cues. J. L. Miller and Dexter (1988) is one notable exception. They found that vowel length had weaker effects on perceived voicing when participants responded quickly in a phoneme judgment task, suggesting that participants made their earliest responses primarily on the basis of the VOT. However, the boundary indicated by these early responses favored more /p/ responses, suggesting that listeners treated the incomplete vowel length as short. Thus, it is not clear which model these data support.

In addition, J. L. Miller and Dexter's (1988) use of a metalinguistic phoneme decision task raises several possibilities. First, the system may be fundamentally buffered, but the act of making an overt phoneme response may force this buffer to be flushed (when it would not normally be during online recognition). Second, it is possible that a representation that is continuous at the phonemic level is buffered before it is available to lexical activation processes; measuring only lower level processes may miss this. Finally, given the uncertainty about whether phoneme representations are used in word recognition (e.g., Gaskell, Quinlan, Tamminen, & Cleland, 2008), it is important to ask when cues affect lexical activation in order to determine whether cues are obligatorily integrated prior to lexical access or whether they can affect activation directly.

The present work used nine-step VOT (voicing) and formant transition (manner) continua, with two vowel lengths, to determine whether early cues are immediately available to lexical access or whether these cues do not play a role in lexical access until later cues become available. We used eye movements to potential referents in order to sample listeners' lexical hypotheses as the signal unfolds over time (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This is a highly sensitive measure of lexical activation (Alloppena, Magnuson, & Tanenhaus, 1998), showing effects of frequency and neighborhood density (Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Dixon, Tanenhaus, & Aslin, 2007) and mapping well onto lexical activation from models like TRACE (McClelland & Elman, 1986). Importantly, eye movements are sensitive to subtle variations in VOT (McMurray, Tanenhaus, & Aslin, 2002), formant transitions (Crosswhite, Masharov,

McDonough, & Tanenhaus, 2008), and vowel duration (Salverda, Dahan, & McQueen, 2003).

METHOD

Participants

Thirty-three undergraduates from the University of Rochester, who were monolingual speakers of English with no known hearing problems, were paid \$10 on each of 2 days.

Materials

Auditory stimuli were synthetic one-syllable words comprising four lexical/phonological contrast sets: two test contrasts (/b/ vs. /p/ and /b/ vs. /w/) and two filler contrasts (/d/ vs. /g/ and /l/ vs. /r/). Each set contained three minimal pairs representing the endpoints of a continuum. Within each continuum, vowel quality was constant. The /b/-/p/ pairs were *beach/peach*, *bees/peas*, and *beak/peak*; the /b/-/w/ pairs were *bell/well*, *bench/wench*, and *belt/welt*. Filler pairs were *deuce/goose*, *dune/goon*, *dew/goo*, *race/lace*, *ray/lay*, and *rake/lake*. Stimuli were synthesized using the KlattWorks (McMurray, 2008) interface for the Klatt (1980) synthesizer. Formant frequencies were modeled after natural tokens produced by a native speaker of English. Other parameters were set with reference to the spectrogram of the recorded word. Parameters for the initial portion of the stimulus were identical across all continua (within a contrast) in order to reduce differences between words and in order to allow analyses to be performed across all continua within a set.

Each stimulus file began with 100 msec of silence. For the voicing continua, words started with a 5-msec release burst at 60 dB. For VOTs of 0 msec, voicing (the AV parameter set to 60 dB) started simultaneously with this burst. To construct each step of the VOT continuum, the onset of AV was delayed in increments of 5 msec from the burst, and 60 dB of aspiration was added, starting from the onset of the burst and ending at the onset of voicing (Figure 1A).

The manner continua contained no release burst, but the amplitude envelope featured a sudden onset (rather than ramping up). Formant transitions were modeled with logistic functions, which are characterized by a smooth transition between steady states. Slope was varied in nine steps, from a steep slope (/b/)—40 Hz/msec for all three formants—to shallow slopes (/w/)—14 Hz/msec (for F2 and F3) and 8 Hz/msec (for F1). As the slope decreased, the midpoint of the function was also delayed (by up to 20 msec for the endpoint /w/; see Figure 1B).

For both continua, vowel durations were based on the duration of the naturally recorded utterance. We created two vowel length conditions by shortening or lengthening the original vowel by 50 msec (see Table 1). Details for how the /l/-/r/ and /d/-/g/ continua were constructed are available in an online supplement.¹ The visual stimuli were 24 canonical pictures depicting each lexical item, edited to remove extraneous content.

Procedure

Participants were first familiarized with the (written) names of the pictures. During testing, four pictures were presented on each trial: two pictures for each experimental pair (e.g., *beach* and *peach*), and two pictures for a filler contrast. Voicing trials were paired with /l/-/r/ fillers. Manner trials were paired with /d/-/g/ fillers. The pairings of words were constant across trials (e.g., *beach/peach* was consistently paired with *lake/rake*) but randomized between participants. Pictures were positioned in the corners of a 19-in. monitor with a central fixation circle. Participants clicked on the circle after it changed color (500 msec after display onset), triggering the auditory stimulus (presented through Sennheiser HD570 headphones). The participants then clicked on the referent with a computer mouse.

Stimuli were presented randomly and repeated five times. Because this resulted in a large number of trials (1,080), the experiment was divided into two 1-h sessions on consecutive days.

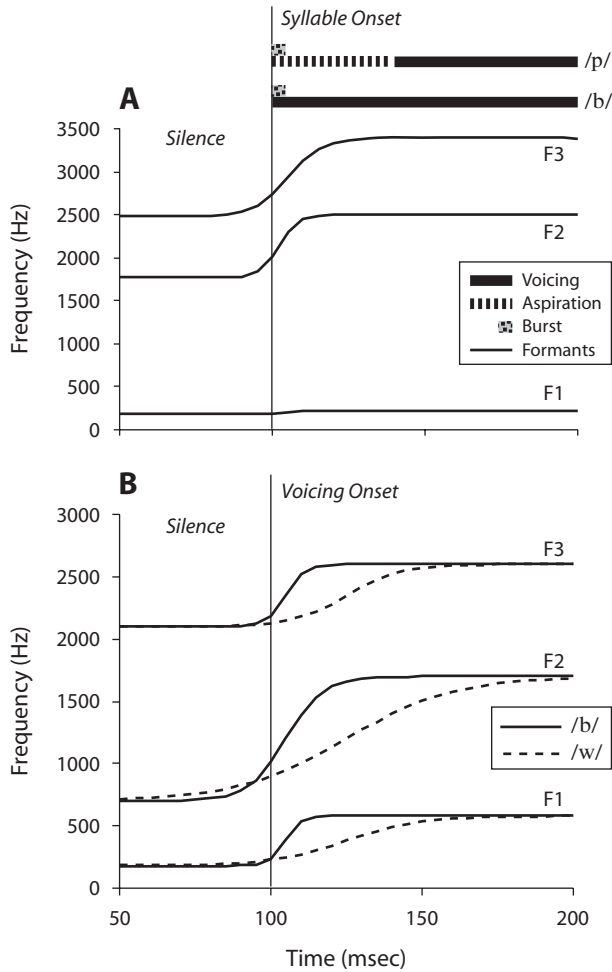


Figure 1. Schematics of the stimuli. (A) For voice onset time (VOT) continua, the frication burst and formant frequencies were kept constant across VOTs, but the relative onset of voicing and aspiration were manipulated. (B) For manner continua, the slopes of the first three formants were manipulated simultaneously.

Eye movements were recorded using an SR EyeLink II eyetracker, sampling at 250 Hz. Eye movements were recorded from the onset of the auditory stimulus until the response. The eye movement record was parsed into saccades and fixations using the default parameter set in the EyeLink software. Saccades that occurred too early to be driven

Table 1
Vowel and Total Length (in Milliseconds) of the Stimuli in Each of the Six Continua

Phonetic Contrast	Word Pair	Vowel Length		Total Length	
		Short	Long	Short	Long
/b/-/p/ (VOT)	<i>beach/peach</i>	105	200	315	415
	<i>bees/peas</i>	145	245	335	435
	<i>beak/peak</i>	105	205	240	340
/b/-/w/ (formant transition)	<i>bell/well</i>	150	250	290	390
	<i>belt/welt</i>	165	265	340	440
	<i>bench/wench</i>	180	280	405	505

Note—For all stimuli, vowel length was measured from the onset of voicing to (1) the offset of voicing, for *beach/peach*, *beak/peak*, and *bench/wench*; (2) the onset of frication, for *bees/peas*; or (3) the end of the second formant transition indicating the /l/, for *belt/welt* and *bell/well*.

by the speech signal were excluded (i.e., those generated during the first 300 msec of the trial: 100 msec of silence at the onset of the auditory stimulus file plus the 200-msec oculomotor planning delay).

Saccades were combined with the subsequent fixation into a *look*, which began at the onset of the saccade and ended at the offset of the fixation. In classifying the object to which looks were directed, the object borders on the screen were extended by 150 pixels in each direction in order to account for noise in the eye track. Adjacent pictures were separated by 480 pixels (horizontal) and 224 pixels (vertical), so there was little chance of a misclassification.

RESULTS

First, we analyzed the mouse-click responses in order to verify that the stimuli produced the desired shifts in category boundaries. Second, and, most important, the fixation data were used to measure the effect of each cue over time. Finally, we assessed whether participants' initial biases reflected an obligatory use of the short portion of the vowel heard at that point.

Voicing

Mouse-click results. Six of the 33 participants categorized end-point stimuli less than 75% correctly and were excluded from the analysis.

As is shown in Figure 2, vowel length shifted the VOT boundary by 8 msec in the expected direction (more voiceless judgments for short vowels). A logistic function was fit to each participant's data for each of the three word pairs and two vowel lengths. The boundaries of these functions were compared in a 2 (vowel length) × 3 (word pair) ANOVA. We found a significant main effect of vowel length [$F(1,26) = 46.7, p = .0001$] but no effect of word pair [$F(2,52) = 2.3, p > .1$]. Vowel length was significant for each continuum [*beach/peach*, $t(26) = 5.8, p = .0001$; *beak/peak*, $t(26) = 5.7, p = .0001$; *bees/peas*, $t(26) = 3.9, p = .001$]. There was a significant interaction [$F(2,52) = 4.6, p = .014$] because the vowel length effect for *bees/peas* was smaller than those for the other continua.

Timing of the effects. Figure 3 shows the likelihood that a look (at any time) was directed to the /b/ or /p/ objects as a function of VOT and vowel length. There are clear effects of both VOT and vowel length throughout the time course of processing.

To evaluate the timing of these effects, we first computed the proportion of fixating each object as a function of time (as is shown in Figure 3) for each participant in each condition. This was smoothed with an 80-msec asymmetrical sawtooth window, in which points 84 msec prior to the current point received a weight of 0 and the weight rose linearly to 1 at the point in question. Thus, any given point was smoothed only by data that occurred before it (so that later cues could not influence the estimate). We then computed a single variable: /b/-/p/ bias—the difference between the likelihood of fixating the voiced or the voiceless competitor (at any given point in time). The value of this variable will be near 0 if participants are equibiased or if the component values are small. To the extent that it is nonzero, it reflects a commitment to either the voiced or the voiceless category.

From this variable, we computed, at each 4-msec time step, a measure of the effect size of VOT and vowel length.

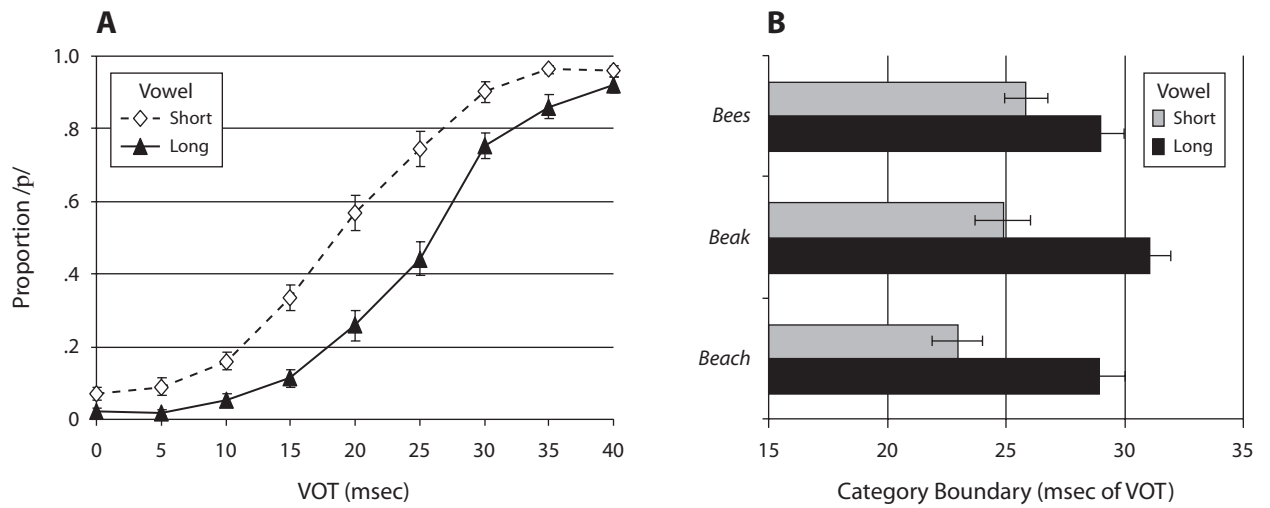


Figure 2. Identification data for voice onset time (VOT) continua. (A) Proportion of /p/ responses as a function of VOT and vowel length averaged across the three voicing continua. (B) Computed category boundaries for each continuum as a function of vowel length.



Figure 3. Effect of voice onset time (VOT) and vowel length on fixations as a function of time. (A) Fixations to voiced targets as a function of VOT and time. (B) Fixations to voiceless targets as a function of VOT. (C) Fixations to voiced targets as a function of vowel length and time. (D) Fixations to voiceless tokens as a function of vowel length.

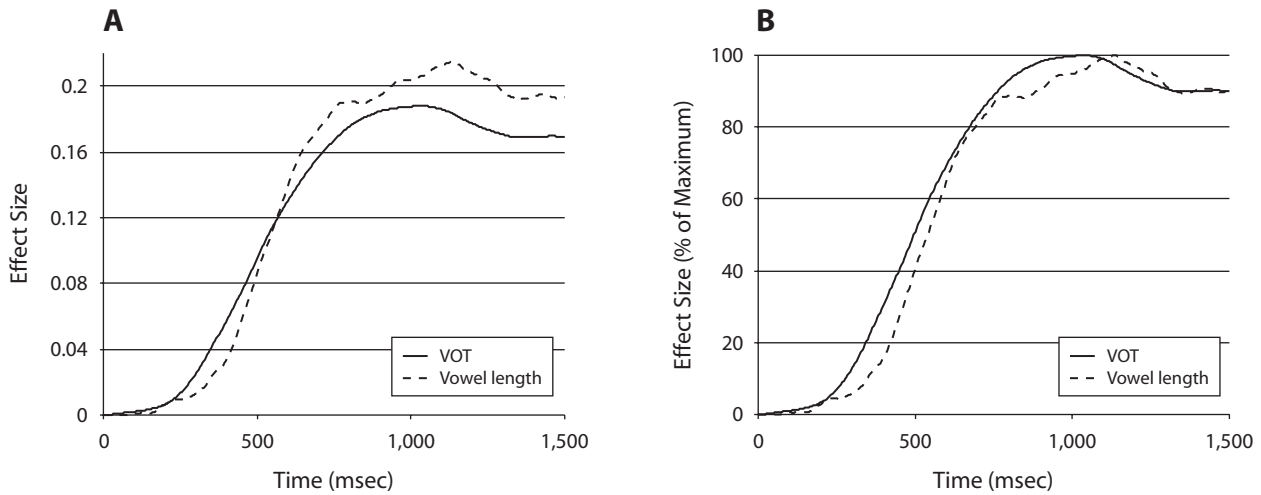


Figure 4. Effect of voice onset time (VOT) and vowel length on fixations as a function of time for voicing continua. 0 msec represents the onset of the stimulus (taking into account oculomotor planning delays). (A) Raw effect sizes. The effect of VOT is the regression slope relating /b/-/p/ bias to VOT at each time step. Vowel length is the difference in /b/-/p/ bias between long and short durations. (B) Effect sizes are rescaled such that 1 is the maximum size (within an effect) and 0 represents no effect. This is the basis of the jackknifing analysis, in which effects were in terms of the percentage of their maximum values.

For VOT, this was the slope of the regression line relating /b/-/p/ bias to VOT. For vowel length, it was the difference in /b/-/p/ bias between long and short vowels. Figure 4 shows the mean effect sizes as a function of time. There is a clear difference in their time course: VOT departs from zero earlier than vowel length does. Since each cue lies along a different scale, in order to compare timing, the onset of a cue was determined as the point at which each effect reached some proportion of its maximum value (e.g., J. O. Miller, Patterson, & Ulrich, 1998). We chose four points: 10%, 15%, 20%, and 30%.

Because fixation data in the visual world paradigm are not typically reliable for individual participants (particularly with small effects), we adapted the jackknife method used by J. O. Miller et al. (1998) for event-related potentials. This technique extracts test statistics from the full data set

averaged across every participant except one, and then repeats this process, excluding each participant in turn. The jackknifed data are then subjected to a statistical analysis in which error terms are adjusted to reflect the fact that each participant contributes $N - 1$ times toward the variance, resulting in a very conservative analysis, particularly for large sample sizes (see Shao & Tu, 1995, for a review).

From the jackknifed data, we first computed the effect size at each time point and, from this, the time at which the effect size in each condition crossed 10%, 15%, 20%, and 30% of its maximum value and remained there for at least 40 msec. This criterion prevented small bumps in the function from driving these estimates.

Fixations were affected by VOT before they were by vowel length. The vowel length and VOT effects were marginally different at the 10% point [$t_{\text{jackknifed}}(26) = 1.87$,

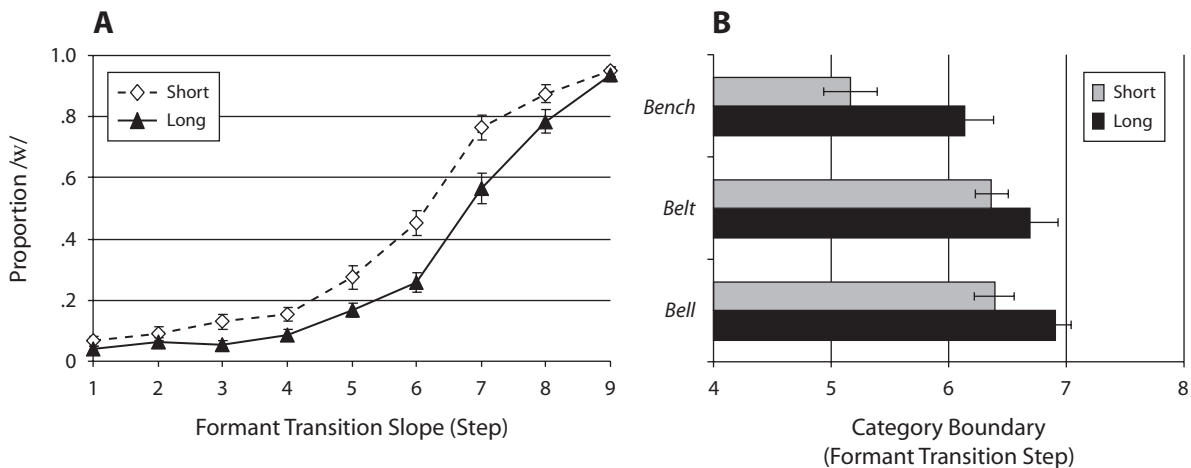


Figure 5. Identification data for formant transition continua. (A) Proportion /w/ responses as a function of formant transition slope and vowel length averaged across the three voicing continua. (B) Computed category boundaries for each continuum as a function of vowel length.

$p = .07$], although the means differed in the expected direction ($M_{VOT} = 279$ msec, $M_{vowel} = 349$ msec). The effects crossed 15%, 20%, and 30% at significantly different times. At 15% of maximum, VOT appeared at 315 msec and vowel length at 395 msec [$t_{jackknifed}(26) = 2.08, p = .047$]. At 20%, VOT appeared at 344 msec and vowel length at 424 msec [$t_{jackknifed}(26) = 3.07, p = .0049$]. And at 30%, VOT appeared at 398 msec and vowel length at 463 msec [$t_{jackknifed}(26) = 2.28, p = .03$].

Initial bias. Our final analysis concerned whether participants treated the early portion of the vowel as short, which would suggest an obligatory use of whatever vowel length information was available at that time (J. L. Miller & Dexter, 1988). If this were the case, early fixations should be biased to the voiceless object, particularly for VOTs near the boundary. To control for the effect of VOT, we examined tokens one, two, and three steps from each participant's boundary and computed the proportion of fixations to the /b/ and /p/ objects generated prior to the end of the short vowels ($M = 296$ msec). At three steps away, participants were significantly biased to /b/ on the voiced side [$t(26) = 3.7, p = .001$] and to /p/ on the voiceless side [$t(26) = 2.7, p = .01$]. However, at steps closer to the boundary, there was no significant bias (all $ps > .2$).

Manner

Mouse-click results. Only 1 participant was excluded using the 75% end-point criterion, leaving 32 participants for analysis.

Figure 5 displays the percentage of approximant (/w/) responses as a function of formant transition slope and vowel length. The boundary is not well centered along this continuum, appearing between Steps 6 and 7. Nonetheless, vowel length shifts the boundary by approximately one step. Logistic functions for each participant for each of the word pairs were fit, and the boundaries of these functions were compared in a 2 (vowel length) \times 3 (word pair) ANOVA. As predicted, we found a significant main effect of vowel length [$F(1,31) = 18.4, p = .0001$]. The word pair effect was unexpectedly significant [$F(2,62) = 13.9, p = .0001$], with the *bench/wench* boundary shifted toward the /b/ end of the continuum, perhaps because of the nasal or the affricate. The interaction was marginally significant [$F(2,62) = 3.0, p = .055$], but all continua had boundary shifts in the expected direction. Paired t tests revealed significant effects of vowel for *bell/well* [Diff. = .52 steps; $t(31) = 3.6, p = .001$] and for *bench/wench* [Diff. = .97; $t(31) = 6.1, p = .001$], but not for *belt/welt* [Diff. = .32; $t(31) = 1.1, p = .2$].

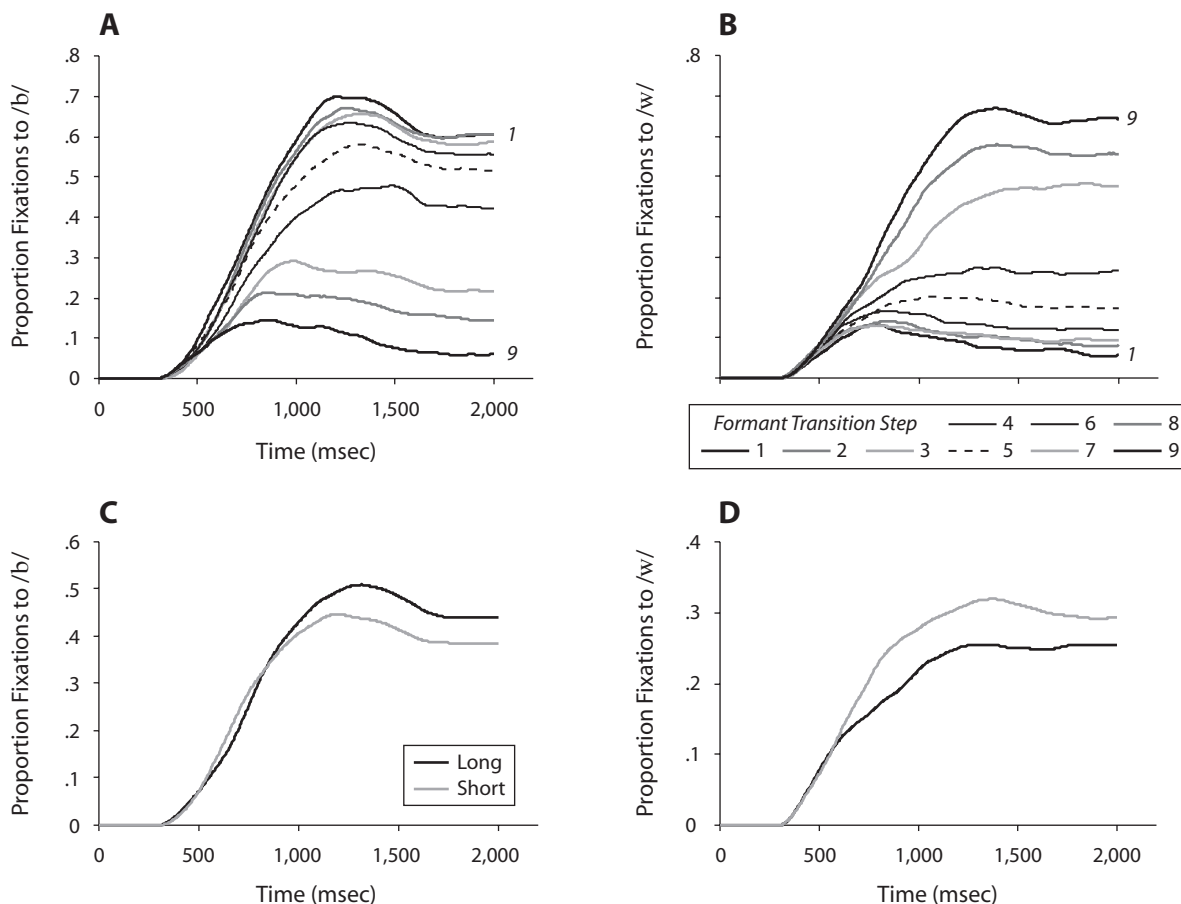


Figure 6. Effect of formant transition and vowel length on fixations as a function of time. (A) Fixations to stop targets as a function of formant transition and time. (B) Fixations to approximant targets as a function of formant transition. (C) Fixations to stop targets as a function of vowel length and time. (D) Fixations to approximant targets as a function of vowel length and time.

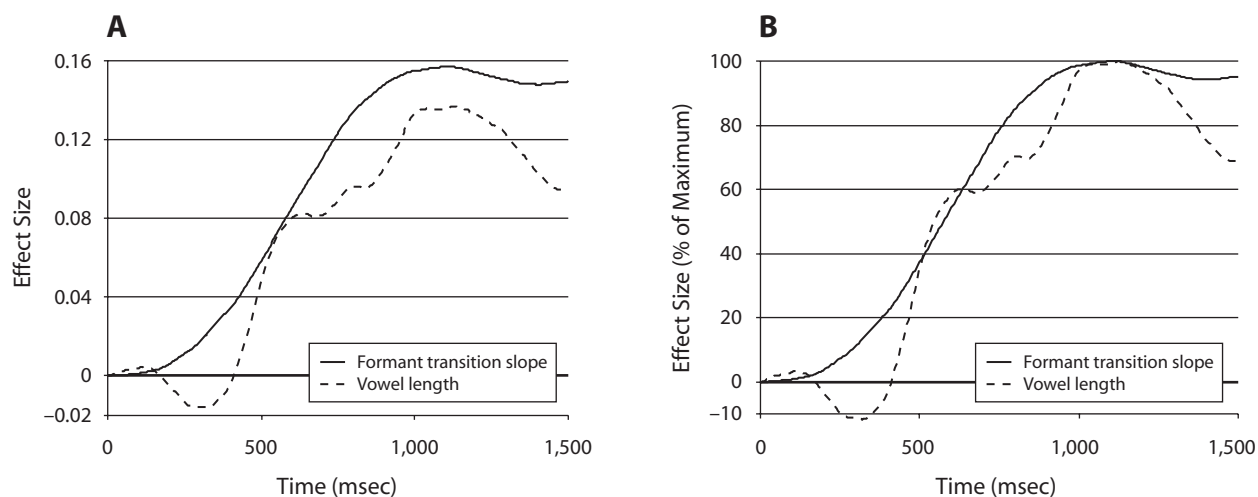


Figure 7. Effect of formant transition slope and vowel length on fixations as a function of time for manner continua. 0 msec represents the onset of the stimulus (taking into account oculomotor planning delays). (A) Raw effect sizes. The formant transition effect is the regression slope relating /b/-/w/ bias to formant transition slope (in step number) at each time step. Vowel length is the difference in /b/-/w/ bias between long and short durations. (B) Effect sizes are rescaled such that 1 is the maximum size (within an effect) and 0 represents no effect. This is the basis of the jackknifing analysis, in which effects were in terms of the percentage of their maximum values.

Timing of the effects. Figure 6 shows the overall pattern of fixations to the stop and approximant competitors as a function of time, formant transition slope, and vowel length. Both experimental factors clearly affected fixations.

Figure 7 shows the effect sizes of the formant transition and vowel length measures on the /b/-/w/ bias measure (looks to /b/ minus looks to /w/) as a function of time after word onset. The effect of formant transition was the regression slope relating step number to /b/-/w/ bias; the effect of vowel length was the difference in /b/-/w/ bias between long and short vowels. The effect of the onset cue (formant transition) departs from zero earlier than vowel length does.

We again used the jackknife procedure to determine when each effect size reliably crossed 10%, 15%, 20%, and 30% of its maximum value and stayed there for 40 msec. The vowel length and formant transition slope effects crossed 10% of their maximum at significantly different times, with formant transition slope appearing at 289 msec and vowel length at 441 msec [$t_{\text{jackknifed}}(31) = 4.94, p = .0001$]. At the 15% point, the average times were also significantly different, with the formant transition slope appearing at 337 msec and vowel length at 454 msec [$t_{\text{jackknifed}}(31) = 3.93, p = .0004$]. Likewise, the 20% point showed a significant difference ($M_{\text{formant}} = 383$ msec, $M_{\text{vowel}} = 467$ msec) [$t_{\text{jackknifed}}(31) = 2.82, p = .008$]. At the 30% point, however, there was no significant difference ($M_{\text{formant}} = 457$ msec, $M_{\text{vowel}} = 489$ msec) [$t_{\text{jackknifed}}(31) = 1.24, p > .2$], reflecting the fact that although it starts later, the effect of vowel length appears to catch up rapidly.

Initial bias. In order to assess whether participants treated the initial portion of the vowel as short (and hence were biased toward /w/), we compared looks to the /b/ and /w/ objects over the first 345 msec (the mean length of the short vowel). We examined three steps on either side

of the boundary. On the stop side, participants were biased toward /b/ at three steps from the boundary [$t(31) = 2.3, p = .03$], marginally biased toward /b/ at two steps [$t(31) = 1.7, p = .1$], and unbiased at one step [$t(31) = 0.9, p > .2$]. On the approximant side, they were biased toward /w/ at three steps from the boundary [$t(29) = 2.8, p = .01$], marginally so at two steps [$t(31) = 1.7, p = .09$], and, again, biased toward /w/ at one step away [$t(31) = 2.4, p = .02$]. Thus, the only evidence for a bias toward /w/ early in processing comes from steps in which the formant transitions were consistent with /w/.

DISCUSSION

Our results provide support for continuous integration of word-initial voicing and manner of articulation based on early consonantal and later vowel-length cues. In both cases, the effect of the onset cue (VOT and formant transition slope) preceded the effect of vowel length, and there was little evidence that the early portion of the vowel was treated as short (biasing toward /w/ or /p/). The system does not appear to wait until both cues are available to make preliminary commitments.

This result offers a partial explanation for why lexical activation is gradiently sensitive to continuous cues like VOT (e.g., Andruski, Blumstein, & Burton, 1994; McMurray et al., 2002). Preserving the continuity of cues like VOT is a fundamental requirement for continuous integration and updating (see also McMurray, Tanenhaus, & Aslin, in press). In contrast, a categorical decision prior to accessing the lexicon would treat a VOT of 25 msec, which is close to the /b/-/p/ boundary and strongly conditioned by vowel length, as equivalent to a VOT of 60 msec, which is less affected by vowel length.

Our results rule out both an encapsulated buffer for the cues that we studied and the notion that integration of cues

is a prerequisite to word recognition. We did not, however, assess other possibilities. Buffering could vary by cue or by contrast, and where buffers are found, they could exist at subcategorical or categorical levels of processing. Such buffers might also vary in their ability to release (cascade) preliminary analyses. For example, Kingston (2005) argued that *integral* cues are integrated at an auditory level and cannot have independent effects on perception, whereas *separable* cues can be weighted according to language-specific phonetic experience.

Although we studied voicing and manner contrasts as representative of the problem, it will be important to consider other cues, as well as cue integration problems that cross word boundaries, such as assimilation and long-distance phonetic dependencies. These results, however, suggest that for at least some cues, word recognition can make immediate, partial commitments without waiting for disambiguating information.

AUTHOR NOTE

This work was supported by NIH Grants DC006537 and DC008089 to B.M., and DC005071 to M.K.T. and R.N.A. The authors thank Dana Subik for assistance with data collection, Steve Luck for suggesting the jackknife analysis, and Gregg Oden for helpful discussion on the nature of cue integration. We particularly thank Joanne Miller for advice, encouragement, and insight during the development of this article. Correspondence concerning this article should be addressed to B. McMurray, Department of Psychology, E11 SSH, University of Iowa, Iowa City, IA 52240 (e-mail: bob-mcmurray@uiowa.edu).

REFERENCES

- ALLEN, J. S., & MILLER, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America*, **106**, 2031-2039.
- ALLOPENNA, P. D., MAGNUSON, J. S., & TANENHAUS, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, **38**, 419-439.
- ANDRUSKI, J. E., BLUMSTEIN, S. E., & BURTON, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, **52**, 163-187.
- CROSSWHITE, K., MASHAROV, M., McDONOUGH, J., & TANENHAUS, M. K. (2008). *Phonetic cues to word length in the online processing of onset-embedded word pairs*. Manuscript submitted for publication.
- DAHAN, D., MAGNUSON, J. S., & TANENHAUS, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, **42**, 317-367.
- FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.
- GASKELL, M. G., QUINLAN, P. T., TAMMINEN, J., & CLELAND, A. A. (2008). The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*, **137**, 282-302.
- GOW, D. W., JR. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, **65**, 575-590.
- KESSINGER, R. H., & BLUMSTEIN, S. E. (1998). Effects of speaking rate on voice onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, **26**, 117-128.
- KINGSTON, J. (2005). Ears to categories: New arguments for autonomy. In S. Frota, M. Vigario, & M. J. Freitas (Eds.), *Prosodies: With special reference to Iberian language* (pp. 177-222). Berlin: Mouton de Gruyter.
- KLATT, D. (1980). Software for a cascade/parallel synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.
- MAGNUSON, J. S., DIXON, J. A., TANENHAUS, M. K., & ASLIN, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, **31**, 133-156.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- McMURRAY, B. (2008). *KlattWorks: A (somewhat) new systematic approach to formant-based speech synthesis for empirical research*. Manuscript in preparation.
- McMURRAY, B., TANENHAUS, M. K., & ASLIN, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, **86**, B33-B42.
- McMURRAY, B., TANENHAUS, M. K., & ASLIN, R. N. (in press). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. *Journal of Memory & Language*.
- MILLER, J. L., & DEXTER, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 369-378.
- MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, **25**, 457-465.
- MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.
- MILLER, J. L., & WAYLAND, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **54**, 205-210.
- MILLER, J. O., PATTERSON, T., & ULRICH, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, **35**, 99-115.
- ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.
- REPP, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81-110.
- SALVERDA, A. P., DAHAN, D., & McQUEEN, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51-89.
- SHAO, J., & TU, D. (1995). *The jackknife and bootstrap*. New York: Springer.
- SHINN, P. C., BLUMSTEIN, S. E., & JONGMAN, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **38**, 397-407.
- SUMMERFIELD, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of the Acoustical Society of America*, **7**, 1074-1095.
- TANENHAUS, M. K., SPIVEY-KNOWLTON, M. J., EBERHARD, K. M., & SEDIVY, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632-1634.
- WAYLAND, S. C., MILLER, J. L., & VOLAITIS, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **95**, 2694-2701.

NOTE

1. Details of the filler stimuli, as well as KlattWorks and WAV files for all stimuli, are available online at www.psychology.uiowa.edu/faculty/mcmurray/publications/mcta_supplement.

(Manuscript received January 7, 2008;
revision accepted for publication June 26, 2008.)