ELSEVIER

Cognition 86 (2002) B33–B42

COGNITION

www.elsevier.com/locate/cognit

Brief article

Gradient effects of within-category phonetic variation on lexical access

Bob McMurray*, Michael K. Tanenhaus, Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

Received 27 March 2002; accepted 16 July 2002

Abstract

In order to determine whether small within-category differences in voice onset time (VOT) affect lexical access, eye movements were monitored as participants indicated which of four pictures was named by spoken stimuli that varied along a 0–40 ms VOT continuum. Within-category differences in VOT resulted in gradient increases in fixations to cross-boundary lexical competitors as VOT approached the category boundary. Thus, fine-grained acoustic/phonetic differences are preserved in patterns of lexical activation for competing lexical candidates and could be used to maximize the efficiency of on-line word recognition. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Speech perception; Spoken word recognition; Categorical perception; Lexical access; Eye movements

1. Introduction

One of the longstanding, and as yet unresolved, puzzles in research on speech perception and spoken word recognition is how listeners rapidly decode highly variable acoustic signals articulated by a speaker's vocal tract into seemingly discrete and invariant phonemes and words. The classic view is that specialized mechanisms extract prototypical categories from the variable input associated with differences in talker, speaking rate, and surrounding phonetic context. These mechanisms essentially strip away and discard the variability to unmask the underlying phonemic or featural prototypes (Stevens, 2002). An alternative view is that the speech recognizer preserves fine-grained acoustic differences and uses systematic covariation to probabilistically weight phonemic and lexical alternatives (Smits, 2001). It is this latter model of speech recognition that we examine in the present report, providing the first evidence that listeners show gradient sensitivity to sub-phonemic variability in consonants during on-line word recognition.

---

* Corresponding author.
E-mail address: mcmurray@bcs.rochester.edu (B. McMurray).

PII: S0010-0277(02)00157-9

Initial investigations of sub-phonemic variation focused on consonant identification and discrimination. These classic studies suggested that some classes of consonants are perceived "categorically", with only a small ambiguous region around the category boundary, and poor discrimination of acoustic differences within a category (e.g. Liberman, Harris, Hoffman, & Griffith, 1957). For example, voiced stop consonants have a short lag between release and the onset of vocal fold vibration (voice onset time or VOT), whereas voiceless stop consonants have a longer lag from release to voicing. Thus, VOT is an important acoustic cue for distinguishing between otherwise identical pairs of consonants (e.g. /b/ and /p/). In a standard two-alternative forced choice task using consonant vowel (CV) syllables, identification functions typically show a step-like change, with VOTs shorter than 15 ms labeled as /b/ and VOTs of 25 ms or longer labeled as /p/. In addition, discrimination performance between sounds with small differences in VOT is relatively poor within a category compared to a difference of the same magnitude when it straddles the category boundary.

Subsequent research has demonstrated that listeners are more sensitive to within-category variation than was originally believed. For example, reaction times in both phoneme identification and discrimination tasks increase as the stimuli approach the category boundary (Pisoni & Tash, 1974). Variations of the original task have also found evidence for non-categorical perception (e.g. Carney, Widin, & Viemeister, 1977; Samuel & Tartter, 1986 for a review). Finally, category goodness judgments differ among stimuli within the same category, with these rating differences becoming large as the stimuli approach the category boundary (Allen & Miller, 2001; Miller, 1997).

Sensitivity to fine-grained differences in VOT takes on added significance because within-phonemic category variation may represent systematic *covariation* with the surrounding phonetic environment (e.g. Fougeron & Keating, 1997; Kessinger & Blumstein, 1998; Repp & Mann, 1982). For example, the voicing category boundary for a word-initial stop consonant can be affected by (among other things) the duration of the following vowel, which itself is affected by the position of a word in an utterance (Fougeron & Keating, 1997) and by speaking rate (Summerfield, 1981). Thus, a word recognition system that is sensitive to such variation will be able to more efficiently integrate contextual information, particularly if this sensitivity is present on-line as the signal unfolds. Probabilistic use of fine-grained acoustic/phonetic variation within consonants could therefore be extremely useful for coping with between- and within-speaker variability, noise in the signal, and context-dependent changes in articulation that are conditioned by later occurring cues to speech rate and phonetic environment.

The word recognition system clearly makes use of sub-phonemic information in vowels (Warren & Marslen-Wilson, 1987), which have long been known to be perceived less categorically than consonants (Fry, Abramson, Eimas, & Liberman, 1962) and may even be processed using separate channels (Schouten & Van Hessen, 1992). For example, lexical access is *impaired* by mismatching coarticulatory information (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Streeter & Nigro, 1979; Whalen, 1991). However, there is less evidence that sub-phonemic variation along typically varying dimensions in consonants (e.g. VOT or formant transitions) affects lexical access.

The most relevant evidence comes from the demonstration by Andruski, Blumstein, and

Burton (1994) that a word with a prototypical VOT activates its lexical representation more strongly than a word with a less prototypical VOT. Andruski et al. modified natural tokens of voiced and voiceless alveolar stop consonants (i.e. /t/-/d/) to produce words beginning with voiceless consonants that were either fully voiceless, 1/3 voiced or 2/3 voiced. The 2/3 voiced tokens primed semantic associates more weakly than the fully voiceless or 1/3 voiced tokens, which did not differ from each other. The VOT for the fully voiceless stimuli was about 80 ms, placing the VOTs for the 2/3 stimuli at about 27 ms, clearly in the /t/ category, but close to the /t/-/d/ category boundary. Thus, the within-category effects come from differences between stimuli that are close to the category boundary (the 2/3 voicing condition) and stimuli that are quite distant from the category boundary (the 1/3 and fully voiceless tokens). While these results demonstrate sub-phonetic/phonemic effects on lexical access, they leave open the crucial question of whether these effects are truly *gradient* or are limited to differences between exemplars that are near and far from the category boundary.

We monitored eye movements as participants used a computer mouse to click on a picture that matched the name of a spoken word. Critical trials contained a token of a word beginning with a bilabial stop consonant synthesized from a nine-step VOT continuum ranging from 0 to 40 ms in 5 ms steps. Displays contained two potential target pictures, e.g. a picture of a bomb and a palm, and two unrelated pictures, one beginning with an /l/ and one beginning with an /ʃ/, as shown in Fig. 1.

We used the eye movement paradigm because the time course of fixations to potential referents in a visual display is remarkably sensitive to the uptake of acoustic/phonetic

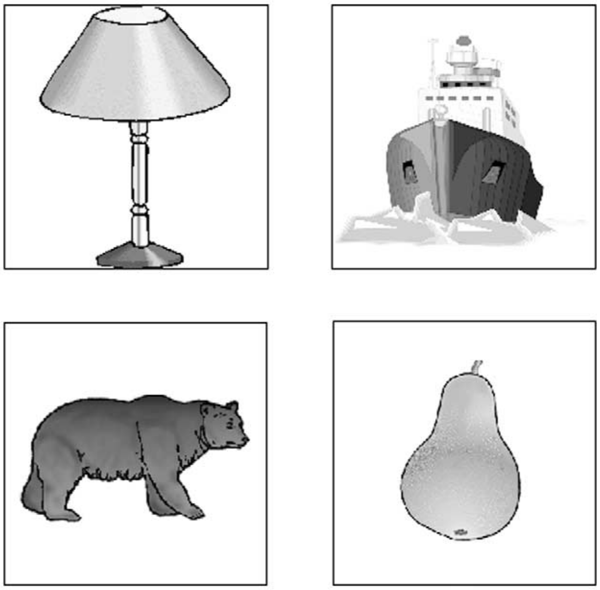

Fig. 1. Sample display of pictures. /b/- and /p/-pictures were consistently paired with each other and with a single /l/- and /ʃ/-picture. Locations of pictures were randomized across trials.

information during spoken language processing (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Moreover, the time course of fixations closely matches predictions generated by a simple quantitative linking hypothesis that maps lexical activation onto proportion of fixations over time (e.g. Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001).[1]

2. Methods

Seventeen University of Rochester undergraduates were recruited from the departmental subject roster, in accordance with the university human subject protocols. They were paid $20.00 for their participation.

Subjects were seated in a quiet room at a Macintosh equipped with a 19 inch monitor and high quality headphones. Eye movements were monitored with an SMI/Eyelink head mounted eye-tracker. The experiment was designed and run using the PsyScope experiment presentation software (Cohen, MacWhinney, Flatt, & Provost, 1993) and the Eyelink PsyExtension (Brenstein, 2000).

Stimuli were synthetic one- and two-syllable words. Twelve of the words comprised minimal pairs from the end-points of a nine-step /b/-/p/ VOT continuum: bomb/palm, bear/pear, beach/peach, butter/putter, bump/pump, and bale/pale. Six synthetic /l/-initial words (lamp, leaf, lip, ladder, leg and lock) and six /ʃ/-initial words (shark, shoe, shirt, shell, ship and sheep) were added as fillers. A small VOT range (0–40 ms) was chosen to allow for a strong test of the hypothesis that listeners are sensitive to small within-category acoustic differences.

All stimuli were synthesized using a preliminary version of the KlattWorks interface to the Klatt (1980) synthesizer (McMurray, 2002).[2] Formant frequencies for the vowels were taken from canonical values described by Ladefoged (1993). All /b/s and /p/s began with a 5 ms burst of frication that was constant for all VOTs. The F1 and F2 formant transitions of the /b/-/p/ stimuli had rising frequencies. Voicing continua for the b/p words were created by cutting back the onset time of the AV (amplitude of voicing) parameter and replacing it with 60 dB of AH (amplitude of frication). All other parameters were kept constant across the continuum. Formant trajectories for the filler stimuli and the final portion of the b/p words were chosen to match as closely as possible spectrographic parameters of natural speech.

Participants heard ten repetitions of each of the nine steps along each of the six b/p-word continua. The same number of repetitions of the /l/- and /ʃ/-words was included to equate the frequency of occurrence of each of the four consonants, yielding a total of 1080 trials.

---

[1] Space does not allow us to address the full set of methodological issues that have been raised and addressed with this paradigm. There is, however, evidence that participants are not comparing the input against stored representations of the names of the pictures activated in working memory (see Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Magnuson, 2001). Perhaps the clearest evidence against this hypothesis is that pictures that share visual similarity with a prototype of the target referent but not name overlap show early competitor effects (Dahan & Tanenhaus, 2002).

[2] Information about the KlattWorks application and scripts for the stimuli are available at http://www.bcs.rochester.edu/~mcmurray/KlattWorks.

Stimuli were presented in random order. Participants completed the experiment over a two-day period with a one-hour session on each day.

The first day began with two blocks of trials to familiarize the participant with the names of the pictures. In the first block, each picture was displayed once with its printed name. In the second block, a picture name was displayed along with four candidate pictures. Participants had to click on the correct picture to advance to the next trial. Two repetitions with each picture as the target resulted in 48 total trials per block.

On both days, the experiment began with the calibration of the eye-tracker using the standard nine-point SMI/Eyelink calibration procedure. Each test trial began with a small blue circle in the middle of the screen and a picture in each quadrant of the screen (e.g. bomb, palm, lamp, and ship). After 500 ms, the blue circle turned red. When the participant clicked on the red circle, a sound-file containing the target word was played. The participant then clicked on the picture named by the word.

The b-, p-, l- and ʃ-pictures appeared in random quadrants of the screen from trial to trial. “B-pictures” were consistently displayed with a single “p-picture” – its competitor (i.e. bomb was always paired with palm). The same l- and ʃ-picture was paired with a given b/p pair so as not to draw the subject’s attention to the relationship between the b- and p-words. For example, a particular participant might consistently see the set: bomb, palm, lamp and ship (although the locations would differ between trials). Pairings of the four word-types were randomized for each participant with the restriction that semantically related pictures such as beach and shell were never displayed together.

Mouse clicks and eye movements were recorded for each trial. The SMI/Eyelink system samples eye position at 250 Hz. The SMI software automatically parses the eye track into three categories: saccades, fixations, and blinks. A look to any of the four pictures was defined from saccade onset to the end of the fixation. Because of inherent drift (over time) in the gaze position reported by the eye-tracker, the borders of the pictures were extended by 100 pixels to capture more fixations. This left a 324-pixel gap between the nearest edges of the defined regions, so there was no danger of assigning an eye movement to the wrong picture.

A separate group of 17 participants used a mouse to click on a P or a B button displayed concurrently on a monitor in response to “pa” and “ba” syllables. These syllables comprised a nine-step 0–40 ms VOT /b/-/p/ continuum, created with the same synthesis procedures used to create the words.

3. Results

Fig. 2 shows the identification function for the mouse click data pooled across all of the b/p words and subjects, as well as the comparable identification function for the “ba” and “pa” syllables. Within the words, the six b/p items showed very similar category boundaries across subjects (17.25 ± 1.33 ms, computed by fitting a logistic function to each subject’s data) and items (17.24 ± 1.24 ms). For the ba/pa identification task, the mean category boundary was 17.5 ± 0.83 ms by subjects. The category boundary did not differ as a function of experiment ($t(26) = 0.36$, $P > 0.1$), although the slope was clearly more step-like for the ba/pa syllables than for the words.

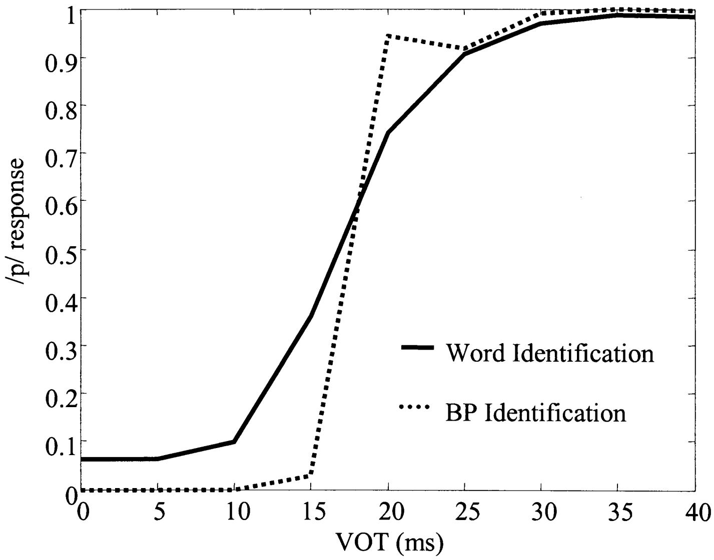

Fig. 2. Identification curves (from mouse clicks) pooled across all subjects for the word and BP identification tasks. Shown is the proportion of trials in which the p-item was selected as a function of VOT.

Fixations were divided into three categories: (1) fixations to the *target* (the picture participants clicked on); (2) fixations to the *related competitor* (the picture with a name that differed from the target name only in voicing); and (3) fixations to *unrelated distracters* (the pictures with names beginning with /l/ or /ʃ/).

For our analyses, we divided the data into two sets: trials in which the target was predominately heard as voiced (/b/: VOT = 0–15 ms) and trials in which the target was predominately heard as voiceless (/p/: VOT = 20–40 ms). In order to conduct a conservative test of the hypothesis that VOT differences within a category have gradient effects on lexical activation, it was important to restrict comparisons to stimuli that were all judged to be exemplars of the same word. Therefore, we eliminated data from **any** trial on which the participant clicked on the competitor picture (i.e. a picture whose name had a VOT value on the opposite side of the category boundary).

Repeated measures ANOVAs (using averaged fixation proportions from 300 to 1900 ms) revealed a significant difference between looks to the competitor and the unrelated items for both the voiced, /b/, and voiceless, /p/, words (/b/: $F(1, 16) = 31.03, P < 0.0001$; /p/: $F(1, 16) = 22.57$, $P < 0.001$). Crucially, VOT interacted with type of target (competitor or unrelated) (/b/: $F(3, 48) = 5.138$, $P = 0.004$; /p/: $F(4, 64) = 10.03$, $P < 0.0001$). Post-hoc tests confirmed that for both /b/ and /p/ this interaction was due to an effect of VOT on looks to the competitor (/b/: $F(3, 48) = 3.30$, $P = 0.028$; /p/: $F(4, 64) = 10.24$, $P < 0.0001$) and not on looks to the unrelated items. All reported analyses showed similar F values and significance levels when words rather than subjects were treated as the random factor.

In order to rule out the possibility that these effects were carried by differences between the end-point stimuli and VOTs near the category boundary, we repeated the analyses only for VOTs that did not abut the 17 ms category boundary (/b/: 0–10 ms; /p/: 25–40 ms). Again there was a main effect of competitor status (/b/: $F(1,16) = 25.74$, $P = 0.0001$; /p/: $F(1,16) = 16.89$, $P = 0.001$) and a significant interaction with VOT (/b/: $F(2,32) = 11.045$, $P < 0.001$; /p/: $F(3,48) = 6.723$, $P < 0.001$).

To test for gradiency, we conducted a one-way analysis on fixations to the competitor. We found significant effects of VOT (/b/: $F(3,48) = 3.73$, $P = 0.017$; /p/: $F(4,64) = 10.16$, $P < 0.0001$) and linear trends (/b/: $F(1,16) = 6.30$, $P = 0.023$; /p/: $F(1,16) = 13.71$, $P = 0.002$). Excluding VOTs near the category boundary (15 and 20 ms) resulted in similar effects of VOT (/b/: $F(2,32) = 4.93$, $P = 0.014$; /p/: $F(3,48) = 6.20$, $P = 0.002$) and linear trends (/b/: $F(1,16) = 8.82$, $P = 0.009$; /p/: $F(1,16) = 9.67$, $P = 0.007$). Fig. 3 presents the mean proportion of fixations to the competitor object for each VOT for the /b/ and /p/ words.

Having established gradient effects on lexical activation with the previous analyses, we examined whether these effects persist beyond the earliest moments of lexical processing. We compared the proportion of times spent looking at competitors in two time windows: from 300 to 1100 and from 1100 to 1900 ms. For each subject, each trial was randomly assigned to one of two categories: "early" or "late". Within each category fixation data were analyzed within only one of these two time windows (300–1100 or 1100–1900 ms). This allowed us to examine two time windows without the possibility of the same fixation contributing to more than one on a given trial, which would violate independence assumptions.

There was a reliable effect of VOT (/b/: $F(3,48) = 3.843$, $P = 0.015$; /p/: $F(4,64) = 5.43$, $P = 0.001$). As VOT approached the category boundary, fixations to

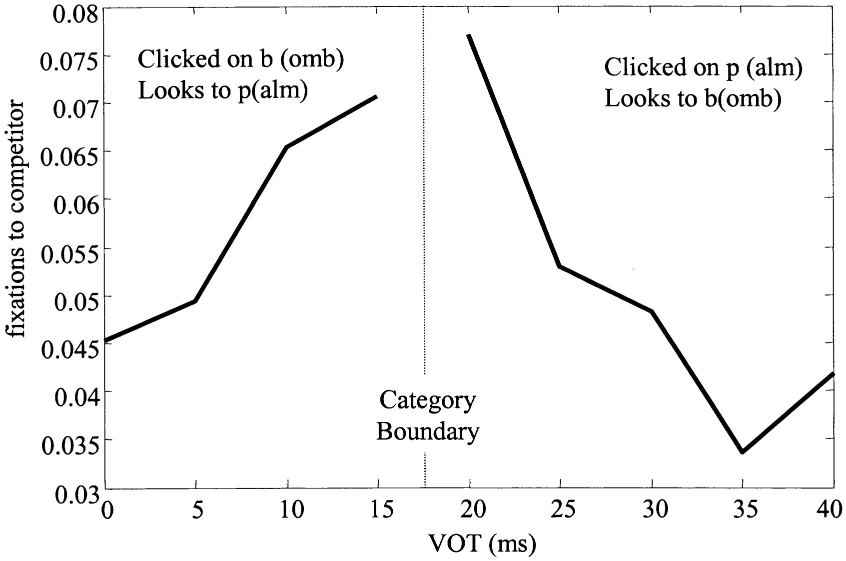

Fig. 3. Mean proportion fixation to the competitor picture as a function of VOT. The left panel displays trials in which the subject responded /b/- (and the competitor was the p-item). The right panel displays trials in which the subject responded /p/- (and the competitor was the b-item). A clear gradient effect of VOT can be seen.

the competitor increased, resulting in a significant linear trend (/b/: $F(1,16) = 6.423$, $P = 0.022$; /p/: $F(1,16) = 8.731$, $P = 0.009$). There was also a reliable effect of time (/b/: $F(1,16) = 17.20$, $P < 0.001$; /p/: $F(1,16) = 29.11$, $P < 0.0001$) such that there were more looks early than late. There was no interaction, however (/b/, /p/: $P > 0.1$), suggesting that although there were fewer looks later, the gradient effect of VOT remained the same.[3]

Similar results were also found when we excluded those stimuli that were closest to the category boundary, restricting analyses to VOTs of 0–10 ms for the /b/ stimuli and 25–40 ms for the /p/ stimuli. There was a significant main effect of VOT (/b/: $F(2,32) = 5.87$, $P = 0.006$; /p/: $F(3,48) = 4.01$, $P = 0.013$) as well as a significant linear trend (/b/: $F(1,16) = 8.06$, $P = 0.012$; /p/: $F(1,16) = 6.64$, $P = 0.020$). Again there was a reliable effect of time (/b/: $F(1,16) = 16.80$, $P = 0.001$; /p/: $F(1,16) = 31.89$, $P < 0.0001$) and no interaction (/b/, /p/: $P > 0.1$).[3]

4. Discussion

This study demonstrates that fine-grained acoustic differences in VOT – differences that have only minimal effects on phoneme identification – nonetheless have significant effects on lexical access.[4] As VOT approaches the category boundary, the activation of a lexical competitor that differs in voicing increases. These gradient effects hold across the full range of VOTs that we tested, suggesting not only that lexical activation functions are much less categorical than phoneme identification functions obtained in standard two-alternative forced choice experiments, but also that such sensitivity approximates the degree of variation in the signal. Gradient effects of VOT persisted for more than a second after word onset, indicating that this information would be available to combine with subsequent information to help resolve temporary phonemic and sub-phonemic ambiguities.

Our results demonstrate that extremely fine-grained acoustic/phonetic differences within phonemic categories are preserved in patterns of lexical activation. To the extent that these fine-grained differences are correlated with the surrounding phonetic environment – a possibility that seems increasingly plausible (Fougeron & Keating, 1997) – they are likely to be exploited by the perceptual system in segmenting and recognizing words from the speech stream. Thus, sub-phonemic information that was presumed to be rapidly

---

[3] A separate analysis for the late time-bin revealed a significant linear trend of VOT for both sides of the continuum using all of the stimuli (/b/: $F(1,16) = 4.80$, $P = 0.044$; /p/: $F(1,16) = 4.51$, $P = 0.05$). When stimuli near the category boundary were excluded this effect remained for /b/ ($F(1,16) = 6.32$, $P = 0.023$) but was not significant for /p/ ($F(1,16) = 1.173$, $P > 0.1$) despite a visually present trend and a reasonable correlation ($R = -0.31$).

[4] It could be argued that lexical access is not involved in this task because subjects adopted a strategy of listening for word initial phonemes given the large number of trials and the uniformity of the competitor set (there was always a b-, p-, l- and ʃ-item). However, we have repeated the experiment using the same stimuli with orthographic buttons (p, b, l and sh) and explicit instructions to listen for word-initial sounds. In this phoneme identification task, the effects of gradiency were markedly reduced (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2002).

discarded by traditional models may instead play an important role in higher-level speech processing.

Acknowledgements

This work was supported by NIH Grants NIDCD DC-005071 and NIDCD T30 DC00035. We thank Dana Subik for help in testing participants and Jim Magnuson for helpful comments.

References

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate. *Perception and Psychophysics*, *63*, 798–810.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Andruski, J., Blumstein, S., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.

Brenstein, R. (2000). The Eyelink PsyExtension. Available at: http://www.psych.uni-potsdam.de/cognitive/eyetracker/eyelink-psyextension.html

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Non categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*, 961–970.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *6*, 84–107.

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.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.

Dahan, D., & Tanenhaus, M. K. (2002). *Activation of conceptual representations during spoken-word recognition*. Paper presented at the 43rd annual meeting of the Psychonomics Society, Kansas City, MO.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*, 171–189.

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.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Ladefoged, P. (1993). *A course in phonetics*. New York: Harcourt Brace.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Magnuson, J. (2001). *The micro-structure of spoken word recognition*. Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, *101*, 653–675.

McMurray, B. (2002). KlattWorks: a [somewhat] new systematic approach to formant-based speech synthesis for empirical research, in preparation.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2002). Two B or not two B: categorical perception in lexical and nonlexical tasks, in preparation.

McQueen, J., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25* (5), 1363–1389.

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes*, *12*, 865–869.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics*, *15*, 285–290.

Repp, B., & Mann, V. A. (1982). Fricative-stop coarticulation: acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *71*, 1562–1567.

Samuel, A., & Tartter, V. (1986). Acoustic-phonetic issues in speech perception. *Annual Review of Anthropology*, *15*, 247–273.

Schouten, M. E., & Van Hessen, A. J. (1992). Modeling phoneme perception: I. Categorical perception. *Journal of the Acoustical Society of America*, *92* (4, Pt 1), 1841–1855.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, *27* (5), 1145–1162.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, *111* (4), 1872–1891.

Streeter, L., & Nigro, G. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, *65*, 1533–1541.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1074–1095.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, *41*, 262–275.

Whalen, D. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, *50*, 351–360.