Published in final edited form as:
Cognition. 2007 November ; 105(2): 466–476.

Effects of prosodically-modulated sub-phonetic variation on lexical competition

Anne Pier Salverda[a,b], **Delphine Dahan**[c], **Michael K. Tanenhaus**[a,d], **Katherine Crosswhite**[e], **Mikhail Masharov**[a], and **Joyce McDonough**[a,d]

[a]*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA* [b]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands* [c]*Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA* [d]*Department of Linguistics, University of Rochester, Rochester, NY, USA* [e]*Department of Linguistics, Rice University, Houston, TX, USA*

Abstract

Eye movements were monitored as participants followed spoken instructions to manipulate one of four objects pictured on a computer screen. Target words occurred in utterance-medial (e.g., *Put the cap next to the square*) or utterance-final position (e.g., *Now click on the cap*). Displays consisted of the target picture (e.g., a cap), a monosyllabic competitor picture (e.g., a cat), a polysyllabic competitor picture (e.g., a captain) and a distractor (e.g., a beaker). The relative proportion of fixations to the two types of competitor pictures changed as a function of the position of the target word in the utterance, demonstrating that lexical competition is modulated by prosodically-conditioned phonetic variation.

1. Introduction

As a spoken word unfolds, lexical candidates that most closely match the input become partially activated and compete most strongly with the target word for recognition (e.g., Luce, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1993). The recognition of a word thus depends, in part, upon which potential lexical candidates provide the closest match to the input. Attempts to characterize the candidates that most significantly compete for recognition with a target word, often referred to collectively as a word's lexical neighborhood, have primarily focused on the nature of the phonemic overlap between potential lexical candidates.

In this article, we demonstrate that the degree to which different lexical candidates compete with one another is modulated by prosodic variation, which systematically affects how a word is phonetically realized within an utterance. We show that naturally occurring prosodic variation affects the relative degree to which polysyllabic and monosyllabic words (e.g., *captain* and *cat*) compete with a monosyllabic target word that shares the same onset and vowel (e.g., *cap*).

The phonetic realization of a word varies with its position in a prosodic domain. Speech sounds preceding a major prosodic boundary, especially vowels, are lengthened (Edwards, Beckman,

Address for correspondence: Anne Pier Salverda, University of Rochester, Department of Brain & Cognitive Sciences, Meliora Hall, Box 270268, Rochester, NY 14627, Tel: +1 585 275 6281, Fax: +1 585 442 9216, E-mail address: asalverda@bcs.rochester.edu.

& Fletcher, 1991; Klatt, 1976; Oller, 1973; Ladd & Campbell, 1991; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Because prosodic boundaries most strongly affect the realization of speech segments in their immediate vicinity, the domain of pre-boundary lengthening does not necessarily correspond to a lexical word: Lengthening affects a monosyllabic word in its entirety, while it primarily affects the final segments of a polysyllabic word.

Listeners have been shown to be sensitive to variations in segmental duration (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). For example, Salverda et al. showed that listeners use segment duration as a probabilistic cue to disambiguate input temporarily consistent with a monosyllabic word, e.g., *cap*, and a polysyllabic word beginning with the same sounds, e.g., *captain*, with longer vowel duration biasing listeners towards the monosyllabic interpretation. The current study extends these results by showing that naturally occurring prosodically-conditioned variation affects listeners' transient consideration of different types of competitors during the recognition of a spoken word. The relative degree to which different types of words compete with one another changes across utterance position.

We examined the recognition of monosyllabic words, such as *cap*, and compared the relative competition from a monosyllabic competitor, e.g., *cat*, with the competition from a polysyllabic competitor in which the target was phonemically embedded at onset, e.g., *captain*. Thus, the phonemic overlap with the spoken word was larger for the polysyllabic competitor than for the monosyllabic competitor. Models of spoken-word recognition where activation strength is primarily determined by phonemic overlap with the spoken input predict that the polysyllabic word will typically be a stronger competitor than the monosyllabic word (McClelland & Elman, 1986; Norris, 1994). By contrast, if the degree of activation of a competitor reflects the match between the spoken input and a representation of the competitor that captures duration (either computed according to the context in which the spoken word occurs or, alternatively, across all past instances of the competitor; e.g. Goldinger, 1998), the match between the initial sounds of the target word and a polysyllabic competitor should be poorer when the target word has undergone lengthening than when it has not. This is because the initial sounds of a polysyllabic word undergo little, if any, lengthening, even before a major prosodic boundary. We induced naturally-occurring variation in the duration of the target word by varying its position in an utterance. The target word appeared in utterance-final position, where it was markedly lengthened (e.g., *Now click on the cap*) or in utterance-medial position, where little lengthening was expected (e.g., *Put the cap next to the square*). If the activation of a competitor varies as a function of the position of the target word and the variation in segmental duration that this position induces, the target word should be less consistent with the polysyllabic competitor than with the monosyllabic competitor in utterance-final position compared to utterance-medial position, where the target word is little lengthened. Thus, if word recognition is sensitive to prosodically induced variation, the relative degree of competition for monosyllabic competitors and polysyllabic competitors should interact with position in a prosodic domain, with stronger competition from monosyllabic competitors in final compared to medial position and the opposite pattern for polysyllabic competitors.

We used the Visual World paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), monitoring listeners' eye gaze as they followed spoken instructions to manipulate (using a computer mouse) one of four pictured objects: the referent, a monosyllabic competitor, a polysyllabic competitor, and a distractor with an unrelated name. This paradigm provides a fine-grained measure of lexical processing over time with a well-defined mapping between the theoretical construct of lexical activation and the observed proportion of fixations to potential referents (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Tanenhaus, Magnuson, Dahan, & Chambers, 2000).

2. Experiment

2.1. Method

2.1.1. Participants—Thirty native speakers of English from the University of Rochester were paid for participation.

2.1.2. Materials—Sixteen triples were constructed, each of which consisted of a monosyllabic target word (e.g., *cap*), a monosyllabic competitor (e.g., *cat*) and a polysyllabic competitor (e.g., *captain*). The choice of items for these triples was highly constrained. Each word had to have an easily pictuerable referent. The logic of the experiment required the polysyllabic competitor word to have the target word phonemically embedded at its onset and the monosyllabic competitor to diverge from the target word at its final segment. It was also important to match the final segment of the target and the monosyllabic competitor for voicing because voicing affects the duration of the vowel preceding that segment. Within the limitations placed by these stimulus constraints, we matched the frequency of the monosyllabic and polysyllabic competitor as closely as possible. However, on average, monosyllabic competitor words were of higher frequency than polysyllabic competitor words (48.4 vs. 13.3 occurrences per million, according to Francis and Kucera, 1982). This pattern reflects a tendency characteristic of the English language in general. Analyses presented in the results section show that frequency differences do not compromise any of the effects of utterance position. Each triple was associated with a phonologically unrelated polysyllabic distractor. In addition to the 16 experimental stimulus sets, listed in the Appendix, 54 filler sets were constructed. Twelve filler trials were included to discourage participants from developing expectations that, in a display comprising pictures with similar names, a monosyllabic word was likely to be the target. These trials had picture names sharing initial segments (e.g., *bull, book*, *bullet*, with one of the monosyllabic words embedded at the onset of the polysyllabic word) with the polysyllabic word as the target. The remaining 42 filler trials consisted of four phonologically unrelated items. A total of 280 pictures [(16+54 trials × 4 pictures)] were selected from various picture databases.

Two instructions were constructed for each experimental trial, varying the position of the referent's name (in utterance-medial position, e.g., *Put the cap next to the square*, or in utterance-final position, e.g., *Now click on the cap*). The same sentence frames were used for the 54 filler sets, with the target word occurring in utterance-medial position in half of these sentences and in utterance-final position in the other half. All sentences were read by a female speaker, a trained phonetician. She was instructed to read the sentences using a natural prosodic phrasing of her choice, as long as this phrasing was consistently used for each type of instruction sentence. Three tokens were recorded for each sentence. The duration of the target word was measured and the token of intermediate duration was used in the experiment. Table 1 presents the average duration of the onset, nucleus and coda of the target words. As expected, the target word was markedly longer in utterance-final position (397 ms) than in utterance-medial position (304 ms), an increase of 31%. The size of this difference is consistent with differences obtained in recordings with naive participants (Crosswhite et al., submitted). Silence of variable duration was inserted at the beginning of the sound file for each utterance to ensure that the same amount of time (i.e., 750 ms) had elapsed from the onset of the sound file to the onset of the target word.

2.1.3. Design—Two lists were constructed by varying which of the two sentences associated with each experimental stimulus set was presented. Within each list, the referent's name occurred in utterance-medial position for half of the experimental items and in utterance-final position for the other half of the items. Three random orders were created for each list. A set of three filler trials was presented at the beginning of the experiment to familiarize participants with the procedure. Fifteen participants were randomly assigned to each list; five were assigned to each randomization.

2.1.4. Procedure—Eye movements were monitored using a head-mounted Applied Sciences Laboratories E5000 eye tracker. A small scene camera aligned with the participant's line of sight provided a continuous recording of the visual scene. Prior to the experiment, the system was calibrated, allowing software to superimpose a participant's point-of-gaze on a HI-8 videotape recording of the scene provided by the scene camera, at a rate of 30 frames per second. Spoken sentences were presented through headphones and simultaneously recorded on the videotape. Participants were first familiarized with each picture, displayed on a computer screen along with its printed name, to ensure its proper identification.

The structure of each trial was as follows. First, a 5×5 grid appeared on the computer screen, with a fixation cross in the center. After a short delay, the experimenter initiated the presentation of the visual display, which was composed of four pictures and four geometric shapes (see Figure 1) and the presentation of the sound file associated with the critical instruction.

2.1.5. Coding procedure—An editing VCR with frame-by-frame controls was used to examine the videotape recording of each participant to establish the location of each fixation (i.e., to the target, the monosyllabic competitor, the polysyllabic competitor, the distractor, or to any other location on the screen). Fixations were coded for each frame on the videotape, starting at the onset of the target word up to and including the time frame when the saccade to the target object that preceded the initiation of a mouse movement to the target object began. When a saccade was in progress, the corresponding time frames were assigned to the target location of the saccade.

2.2. Results and discussion

Of the 480 trials, 8 were lost because of technical failure, track loss or because participants moved the target picture without fixating it. Another 8 trials were discarded because participants selected the competitor picture (1 monosyllabic and 2 polysyllabic in medial condition and 4 monosyllabic and 1 polysyllabic in final condition). For each 33-ms time interval starting at target-word onset, we computed fixation proportions to each picture in the display, for each condition and for each participant or item. Fixation proportions were averaged across participants and items, and, at the suggestion of reviewers, arcsine transformed to improve normality and homoscedasticity of the data.

Proportions of fixations over time to the target, the monosyllabic competitor, the polysyllabic competitor and the distractor are presented in Figures 2 (medial position) and 3 (final position). Throughout most of the time interval displayed on the graphs, the polysyllabic competitor attracted more fixations than the monosyllabic competitor in medial position, while the opposite was observed in final position.[1] To statistically confirm this pattern, we computed the averaged proportion of fixations to the pictures over the time interval from 200 to 1000 ms after target-word onset. The onset of this time interval takes into account the time it takes to program and launch a saccade with a multi-target display (Hallett, 1986). By 1000 ms, fixations to targets have typically reached asymptote in similar studies. The proportion of fixations to monosyllabic and polysyllabic competitors differed across utterance position. In medial position, the probability of fixating the polysyllabic competitor was greater than that of fixating

[1]Although a precise analysis of the time course of competitor fixations is beyond the scope of this study, an aspect related to the time course of fixations deserves further examination: the apparent early advantage for the polysyllabic distractor in medial condition. We have no explanation for this distractor bias, in particular because no comparable effect was observed in final condition. However, this early (albeit non-significant) bias may explain why in medial condition, fixations to the target and competitors appear to be delayed compared to standard cohort effects observed with this paradigm.

the monosyllabic competitor (17% vs 14%,) whereas the reverse was true in final position (13% vs 18%). A two-way ANOVA on the fixation proportions revealed a significant Position by Type of Competitor interaction ($F_1(1,29) = 8.3$, $p < .01$; $F_2(1,15) = 9.6$, $p < .01$). Planned *t*-tests tested the effect of position on the probability of fixating the monosyllabic and polysyllabic competitors separately. The effect of position was significant for monosyllabic competitors (14% in medial position vs. 18% in final position; $t_1(29) = 2.6$, $p < .01$, $t_2(15) = 2.1$, $p < .05$) and for polysyllabic competitors (17% in medial position vs. 13% in final position; $t_1(29) = 2.2$, $p < .05$, $t_2(15) = 2.6$, $p < .05$). These analyses suggest that monosyllabic competitors competed for recognition more strongly in final position than medial position, whereas polysyllabic competitors competed for recognition more strongly in medial than final position.

Because monosyllabic competitors were on average of higher frequency than polysyllabic competitors, it is possible that the increase in activation for monosyllabic competitors in final position compared to medial position reflects a stronger influence of frequency on competitor activation in final position, rather than the impact of a greater acoustic/phonetic match between the spoken word and the representation of a monosyllabic word in final vs. medial position. To address this concern, we conducted three types of analysis. First, we established that a numerically comparable pattern was present on the subset of five triples where the polysyllabic competitor was more frequent than the monosyllabic competitor: Polysyllabic competitors received more fixations than monosyllabic competitors in medial position (21% vs 14%) and fewer fixations in final position (15% vs 21%). Second, the interaction between Position and Type of Competitor remained significant in an ANCOVA on the item analysis using the difference in log frequency between the monosyllabic and polysyllabic competitors as a covariate ($F_2(1,14) = 8.6$, $p < .05$).

Third, we compared linear regression models with and without frequency and word type. In a model with Position and Type of Competitor as main effects and the Type by Position and Frequency by Position interactions, the Type by Position interaction accounted for a significant proportion of the variance ($t(59) = 2.1$, $p < .05$), whereas the Frequency by Position interaction did not ($t(59) < 1$). Moreover, this model was significantly better than the same model with the Type by Position interaction removed (comparison of model residuals: $F(1,59) = 4.3$, $p < .05$). In the latter model, the Frequency by Position interaction did not account for a significant proportion of the variance ($t(59) < 1$) [2].

While compelling, the preceding analyses violate one of the assumptions underlying the the analysis of variance (i.e., the independence of observations). We also report results from a ratio analysis that mitigates this issue. For each participant and each item, the ratio between the proportion of fixations to the monosyllabic competitor and the sum of the proportion of fixations to the monosyllabic and polysyllabic competitor over the 200–1000 ms interval was computed. This ratio expresses the relative degree to which the monosyllabic competitor was fixated. Ratio values greater than .50 indicate the tendency to fixate the monosyllabic competitor more than the polysyllabic competitor; values smaller than .50 indicate the reverse tendency. In medial position, the probability of fixating the monosyllabic competitor over that of fixating either of the two competitors was .46. In final position, the probability was .61, a significant increase ($t_1(29) = 3.4$, $p < .005$; $t_2(15) = 3.0$, $p < .005$). This confirms that the position of the target word affected the relative degree to which monosyllabic and polysyllabic competitors were considered for recognition.

[2] In response to a reviewer's concern, we conducted analyses on shorter time windows. The results of these analyses also argue against any significant influence of lexical frequency on the tendency to fixate the monosyllabic competitor more than its polysyllabic counterpart in utterance-final condition. In particular, during the time interval in which fixation proportions to the monosyllabic competitor increase relative to fixation proportions to the polysyllabic competitor, the size of this effect is not correlated with the difference in log frequency between the competitors, and the proportion of fixations to the monosyllabic competitor is not correlated with its log frequency.

3. General Discussion

Variation in the realization of a spoken word due to its position in an utterance influences the degree to which different types of competitors are involved in the competition process. When a monosyllabic word occurs in utterance-final position, competition from monosyllabic competitors *increases* (compared to processing of the same word in utterance-medial position), while competition from polysyllabic competitors *decreases*.

Our results demonstrate that naturally-occurring phonetic variation conditioned by constituent-level prosodic structure can play a central role in the evaluation of lexical candidates. This finding complements previous demonstrations of sensitivity to artificially created variation (Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002; Salverda et al., 2003). However, the primary contribution of the present research is that it demonstrates that prosodically conditioned variation in the realization of words in continuous speech can act to modulate the lexical competition process dynamically by having a different impact on the evaluation of different types of candidate words (see Magnuson, Dixon, Tanenhaus, and Aslin, in press, for converging evidence). The importance of prosodic effects can be appreciated by noting that in utterance final position, where prosodic effects are strongest, “prosodically matching” competitors competed more strongly than “prosodically mismatching” competitors with greater initial phonemic overlap. Our results appear difficult to reconcile with any model of spoken-word recognition in which lexically contrastive information is represented exclusively along phonemic dimensions (see Salverda et al., 2003, for discussion of how existing models might be extended to accommodate our results). It also suggests that metrics of processing difficulty associated with the composite effects of competitors (e.g. neighborhood density) will need to take into account how prosodically-conditioned variation affects potential competitors.

Acknowledgements

Supported by HD DC005071 to M.K.T. and NSF grant 0433567 and HD49742 to D.D. We thank three anonymous reviewers for their insightful comments, Jessica Maye for recording the stimuli and Florian Jaeger for statistical advice.

References

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;38:419–439.

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

Cooper RM. The control of eye fixation by the meaning of spoken language. A new methodology for the real time investigation of speech perception, memory, and language processing. Cognitive Psychology 1974;6:84–107.

Crosswhite K, McDonough J, Masharov M, Tanenhaus MK. Phonetic cues to word length in the online processing of onset embedded word pairs. Journal of Phonetics. submitted

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

Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes 2001;16:507–534.

Davis MH, Marslen-Wilson WD, Gaskell MG. Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance 2002;28:218–244.

Edwards J, Beckman ME, Fletcher J. The articulatory kinematics of final lengthening. Journal of the Acoustical Society of America 1991;89:369–382. [PubMed: 2002175]

Francis, WN.; Kucera, H. Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton Mifflin; 1982.

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

Hallett, PE. Eye movements. In: Boff, KR.; Kaufman, L.; Thomas, JP., editors. Handbook of perception and human performance. New York: Wiley; 1986. p. 10.1-10.112.

Klatt DH. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 1976;59:1208–1221. [PubMed: 956516]

Ladd, DR.; Campbell, WN. Proceedings of the XII th International Congress of Phonetic Sciences. Aix-en-Provence; France: 1991. Theories of prosodic structure: evidence from syllable duration; p. 290-293.

Luce, PA. PhD dissertation, Indiana University. Technical Report No. 6. Speech Research Laboratory, Department of Psychology; Indiana University: 1986. Neighborhoods of words in the mental lexicon.

Luce PA, Pisoni DB. Recognizing spoken words: The Neighborhood Activation Model. Ear and Hearing 1998;19:1–36. [PubMed: 9504270]

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word learning and recognition: studies with artificial lexicons. Journal of Experimental Psychology: General 2003;132:202–227. [PubMed: 12825637]

Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. Cognitive Science. in press

Marslen-Wilson, WD. Issues of process and representation in lexical access. In: Altmann, GTM.; Shillcock, R., editors. Cognitive models of speech processing: The second Sperlonga meeting. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1993. p. 187-210.

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

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

Norris D. Shortlist: a connectionist model of continuous speech recognition. Cognition 1994;52:189–234.

Oller DK. The effect of position in utterance on speech segment duration in English. Journal of the Acoustical Society of America 1973;54:1235–1247. [PubMed: 4765808]

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

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

Tanenhaus MK, Magnuson JS, Dahan D, Chambers C. Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. Journal of Psycholinguistic Research 2000;29:557–5580. [PubMed: 11196063]

Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America 1992;91:1707–1717. [PubMed: 1564206]

Appendix: Stimulus sets

The first member of a pair of distractors marked with an asterisk was replaced with the second member of that pair after 6 participants had been tested. These distractors received an unusually high proportion of fixations, presumably because the initial sounds of their names were too similar to the initial sounds of the target word.

Target	**Monosyllabic competitor**	**Polysyllabic competitor**	**Distractor**
beak	beet	beaker	whistle
bell	bed	bellows	scissors
bowl	bone	boulder	fountain
bug	bud	buggy	shovel
cap	cat	captain	guitar*/beaker
carp	cart	carpet	ladder
doll	dog	dolphin	magnet
leaf	leash	leaflet	cigar
neck	net	nectarine	letter
pad	pan	paddle	bucket
pick	pit	pickle	ribbon
rack	rat	racket	garlic
robe	road	robot	table
tack	tap	taxi	dagger*/lemon
track	trap	tractor	lighter
well	web	welder	feather

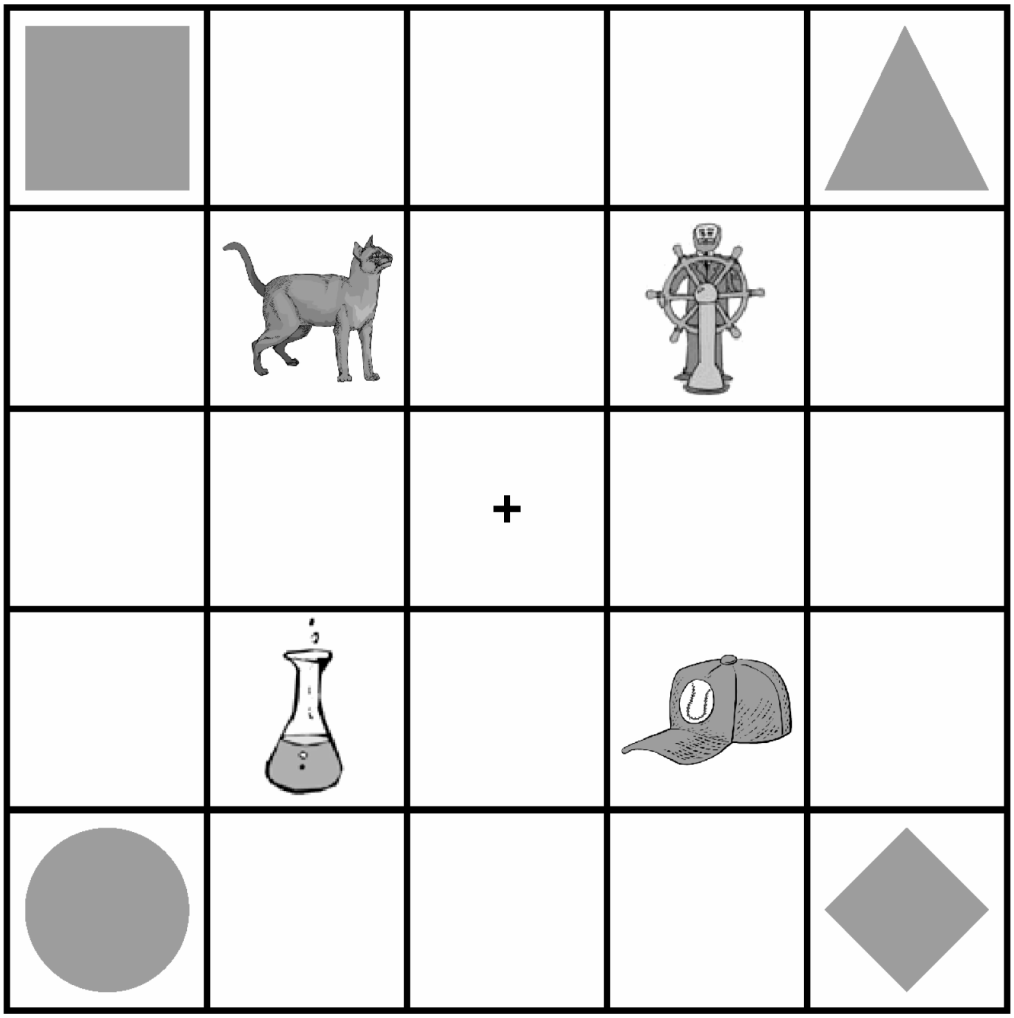

Figure 1.
Example of a visual display. Picture locations were randomized across trials.

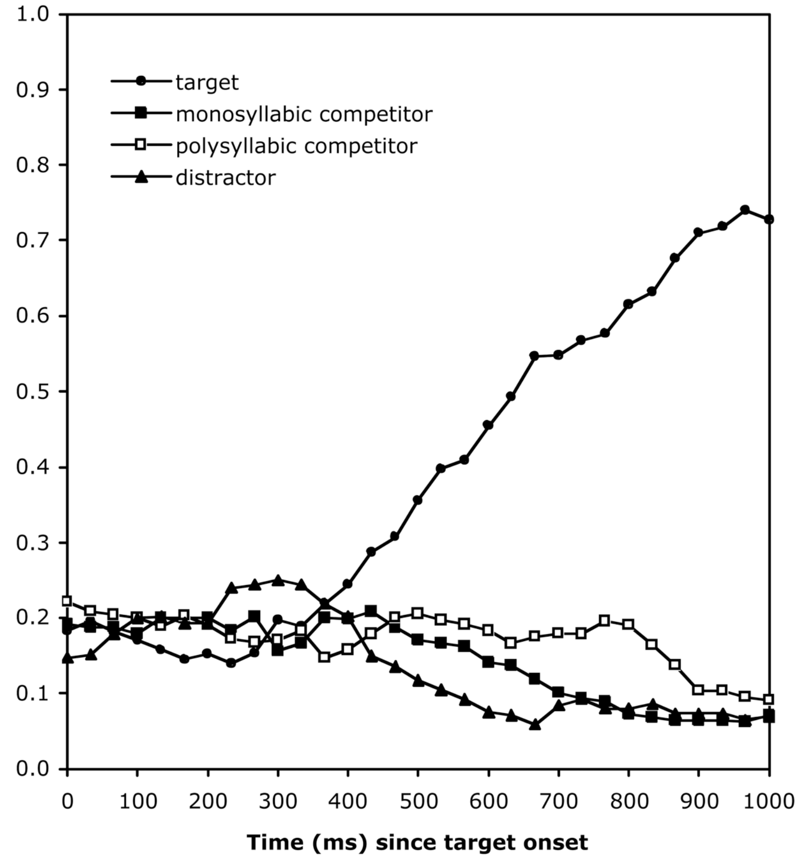

Figure 2.
Proportion of fixations over time to the target, the monosyllabic competitor, the polysyllabic competitor and the distractor, in utterance-medial condition.

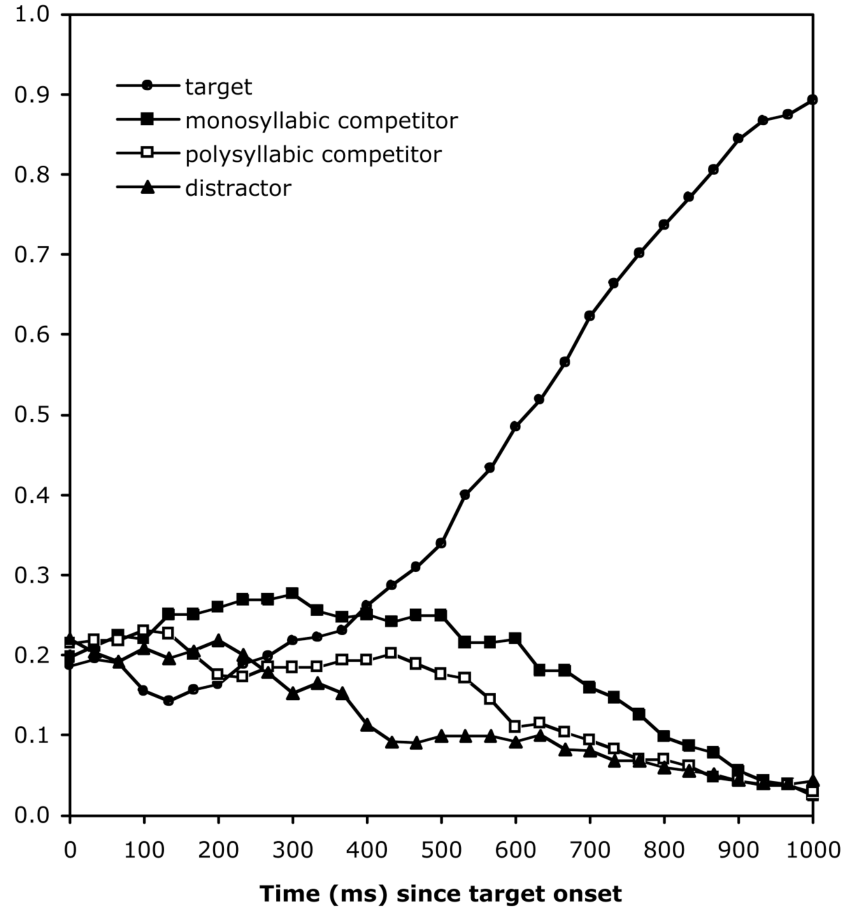

Figure 3.

Proportion of fixations over time to the target, the monosyllabic competitor, the polysyllabic competitor and the distractor, in utterance-final condition.

Table 1

Mean duration (in ms) of the segments of the monosyllabic target word in the utterance-medial (e.g., *Put the cap next to the square*) and utterance-final condition (e.g., *Now click on the cap*).

	Utterance-medial	**Utterance-final**	**Difference**	**Lengthening**
Onset	67	66	−1	−1%
Nucleus	137	165	28	20%
Coda	100	166	66	66%
Total	*304*	*397*	*93*	***31%***