The salience of consonants and vowels in learning an artificial lexicon: The effects of noise

Sarah C. Creel, Richard N. Aslin, & Michael K. Tanenhaus

Department of Brain and Cognitive Sciences
University of Rochester, Rochester, NY, US
screel@bcs.rochester.edu

Abstract

We examined the effects of acoustic noise on the relative weightings of different segment types during lexical learning and lexical retrieval. We designed an artificial lexicon consisting of 16 CVCV words, learned as names for black-and-white shapes. For each word (*dabo*), another word shared the same consonants (*dubei*) and a third had the same vowels (*gapo*). In a 4AFC task, 4 pictures were presented and the name of one picture was spoken. The participant was asked to click on that picture. On each trial, one picture had a name with the same vowels or same consonants as the target picture. Whether words were learned and tested with or without noise (3 dB SNR), overall error rates were roughly equivalent, but vowel-matched errors were much higher in the noise conditions than in the clear conditions. These results support a mechanism of adaptive plasticity whereby acoustic/phonetic cues are re-weighted to maximize the intelligibility of words during lexical learning and lexical retrieval.

1. Introduction

Current models of spoken word recognition assume that all word candidates matching the unfolding acoustic-phonetic input at a given time point are activated and compete for recognition with the best matching candidate emerging as the winner. [1,2,3]. Words sharing onsets ("cohort" words) are the strongest competitors early on, while later-overlapping competitors, such as rhymes, may show some activation at later time points as they begin to match the word. Various factors modulate the strength of these competitors, including prosody [4], segmentation context [5], the competitors' frequencies relative to the target word's frequency [6,7], and category goodness [8].

Magnuson et al. [7] examined the effect of competitors using an artificial lexicon of names for novel black-and-white shapes. The lexicon had been carefully constructed so that each word (*pibo*) had a cohort competitor (*pibu*) and a rhyme competitor (*dibo*). Some words occurred frequently in training, while others occurred infrequently. As well, frequency of target words was crossed with frequency of the word's competitors. Using a head-mounted eye tracker to monitor participants' visual fixations to referents of target and competitor items in a 4AFC task, the authors found strong competition from cohort competitors, and competition from rhyme competitors that waned over the course of lexical learning. In addition, high-frequency words were activated faster than low-frequency words, and high-frequency words were stronger competitors than low-frequency words.

Creel et al. [9] extended this paradigm, focusing on labeling errors rather than eye-fixation proportions during the earliest phase of lexical learning. They found not only that cohort errors exceeded rhyme errors, but also that words matched on syllable-initial consonants were more often confused with one another than words matched on vowels. Thus, words disambiguated by their consonants were less confusable, suggesting that consonants contribute more information to word identity.

There are several possible sources of this consonant bias in word learning. Consonants vary less than vowels across English dialects and across talkers, and consonants have somewhat sharper category distributions than vowels. Moreover, syllable-initial consonants provide temporally earlier information for word recognition than vowels. But how fixed is this consonant bias in lexical access? Creel et al. [9] found that listeners strongly confuse vowel-matched words (*isfim*, *ikfip*; see [9], Experiment 6) when syllable-initial consonants were uninformative. However, error rates were uncharacteristically high in this experiment, suggesting that all words were being confused on the single redundant consonant /f/. Thus, although same-vowel confusions are evident, some bias to identify words by their syllable-initial consonants still exists even when this consonant information is uninformative.

However, the speech-in-noise literature tells a different story. When embedded in noise, segmental confusions increase asymmetrically, with consonant identification more detrimentally affected than vowel identification, largely because vowels are more acoustically robust [10,11]. While not true of all noise types, this is the case for white noise [11] and multitalker babble [10], two prevalent noise sources in modern industrialized society. Under such conditions, a fixed dependence on less-reliable consonantal information during lexical learning and lexical access would be disadvantageous. An inflexible consonant bias would predict continued reliance on consonants, with greater rates of confusion. In contrast, a more exemplar-based theory of lexical learning and access [12] would predict that lexical access would opportunistically take advantage of the information that is reliably present in the input and stored in the lexicon. Such a system would weight vowels more strongly when in the presence of noise.

Here, we test these predictions with an artificial lexicon, where each word has a consonant-matched competitor and a vowel-matched competitor. The lexicon is learned with or without noise, and then participants identify words with or without noise. If the consonant bias in lexical access is immutable, noise will only serve to increase errors. However, if the consonant bias stems from the fact that consonant information is more reliable across accumulated traces, then noise should create more same-vowel confusions by making consonant information less reliable.

2. Method

2.1. Participants

Participants were 92 University of Rochester undergraduates who reported no history of hearing problems. They were paid $10 each for an experiment that lasted 20-30 min.

2.2. Stimuli

2.2.1. Words.

There were 16 words in the artificial lexicon, constructed from the consonants b, d, g, p, and the vowels a, eɪ, o, and u (i was eliminated because many two-syllable English words end in i: e.g. *puppy*, *happy*, *body*). Each word had one other word which shared its consonants, and one other word that shared its vowels, but not both. Four different random assignments of segments to word positions created four different lexicons, to minimize idiosyncratic word-specific effects.

2.2.2. Visual stimuli.

The 16 words were used as labels for a set of 16 unfamiliar black-and-white drawings created in AppleWorks Paint for an earlier study [9]. Six different random assignments of labels to pictures were used.

2.3. Procedure

2.3.1. Training.

PsyScope experimental presentation software [13] was used to present the stimuli. On each training trial, a picture appeared in the center of the screen, and after 750 ms, its name was spoken. The participant then mouse-clicked the picture to proceed to the next trial. Each word and its paired picture were presented 24 times each, in the center of the screen (200 x 200 pixels), for a total of 384 learning trials.

2.3.2. Test.

In testing, four pictures at a time appeared: above, below, to the left, and to the right of center, as depicted in Figure 1. After 100 ms, the name of one of them was spoken, and then the word "Next" appeared in the center of the screen. Participants clicked on what they thought was the appropriate picture. The mouse-click caused the four pictures to disappear and participants clicked on the word "Next" to proceed to the next test trial. Each picture appeared four times as a target: twice with a same-consonant (same-C) competitor's referent present, and twice with a same-vowel (same-V) competitor's referent present, for a total of 64 test trials. The other two pictures in a trial were phonologically unrelated to the target, overlapping in none of the four segment positions (CVCV). We did not include filler trials, because we wished to obtain maximum statistical power and because we deemed the similarity manipulation sufficiently subtle to evade participants' notice.

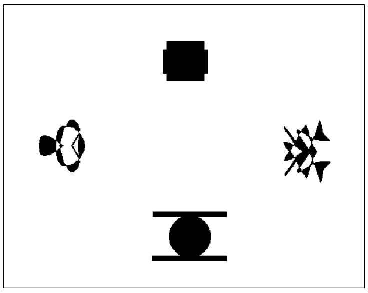

Figure 1: Test display. On each trial, participants selected one of four items as the target when a word was spoken.

3. Results

3.1. Error data

Eleven participants who did not score 40% correct or better were eliminated from the analyses, as it was impossible to determine that such scores resulted from anything but chance performance (clear-clear [CC]: 1; clear-noise [CN]: 4; noise-clear [NC]: 5; noise-noise [NN]: 1). We calculated the percentage of erroneous selections of same-C (or same-V) competitor pictures, and the percentage of erroneous selections of phonologically unrelated distractor pictures on the same trials. Unrelated-picture errors were divided in half to correct for the fact that there were twice as many distractors in a trial (2) as competitors (1). More errors to same-C (or same-V) competitor pictures than to unrelated ones would indicate that significant confusions were taking place.

As depicted in Table 1, there were more errors overall in the mixed conditions (CN, NC) than in the homogenous conditions (CC, NN), which did not differ in overall error rates. However, even though these two conditions had equivalent error rates, confusion patterns (Figure 2) differed markedly: same-C confusions were roughly similar, but same-V confusions were far greater in the noise condition. The CN and NC conditions showed strong confusions of both same-C and same-V words, and the NC condition showed elevated error rates.

Condition	n	Percent Correct	Standard Error
CC	21	84.8%	3.9%
CN	18	77.0%	3.4%
NC	19	71.4%	3.8%
NN	23	83.4%	3.5%

Table 1: Overall percentages correct in each condition.

Statistical analysis confirmed these observations. First, we compared overall error rates between conditions with a repeated-measures ANOVA, where the factors were Exposure Noise (clear, noise) and Test Noise (clear, noise).

There were no main effects, but an interaction ($\mathbf{F}(1,77) = 7.2$, $\mathbf{p} < .009$) such that, while errors in the clear exposure-clear test (CC) case did not differ from the noise-exposure noise-test (NN) case ($\mathbf{t}(42) = .26$, $\mathbf{p} = .79$), errors were lower than in the noise-clear (NC) case (CC: $\mathbf{t}(38) = 2.45$, $\mathbf{p} = .02$; NN: $\mathbf{t}(40) = 2.32$, $\mathbf{p} = .03$). Clear-noise (CN) error rates differed from none of the others.

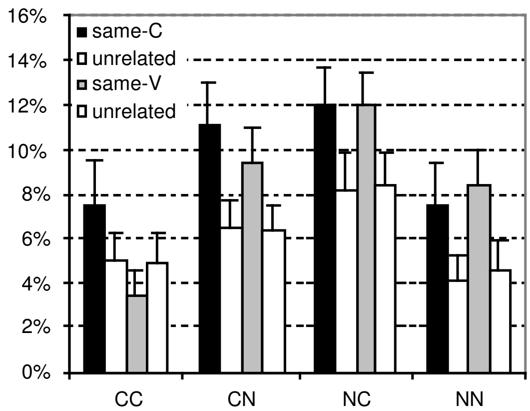

Figure 2: Error data in each condition. Black bars are same-C errors, gray bars are same-V errors, and white bars are unrelated distractor errors. Error bars are standard errors.

A planned ANOVA on same-C confusions, with Exposure Noise (clear, noise), Test Noise (clear, noise), and Error Type (similar, phonologically unrelated) as factors, indicated that consonant confusions were present in all Exposure-Test pairings: there was an effect of Error Type ($\mathbf{F}(1,77) = 19.96$, $\mathbf{p} < .0001$), such that same-C errors exceeded unrelated errors. There was also an interaction of Exposure x Test ($\mathbf{F}(1,77) = 5.58$, $\mathbf{p} = .02$), resulting from the aforementioned higher error rates in the NC case, but the individual Error Type effects did not differ from one another ($\mathbf{p} \geq .37$). A similar ANOVA on same-V confusions showed an asymmetry between conditions; in addition to the effect of Error Type ($\mathbf{F}(1,77) = 10.85$, $p < .002$), and the Exposure x Test interaction ($\mathbf{F}(1,77) = 9.21$, $\mathbf{p} = .003$), there was an Error Type x Exposure ($\mathbf{F}(1,77) = 4.22$, $\mathbf{p} = .04$) interaction. This stemmed from the same-V Error Type effect being smaller (actually negative) in the CC case and greater in the other cases ($\mathbf{p} \leq .02$ in all three comparisons).

While same-C confusions were greater than same-V confusions in the clear conditions, as expected, same-C and same-V competition effects were both evident in the noise conditions. This was true even though overall errors did not differ, suggesting that learners in noise were not learning worse but differently. In the CN and NC cases, both same-C and same-V errors were present, demonstrating effects of noise on both lexical learning and lexical access.

4. Discussion

Several interesting pieces of evidence have emerged from this study. First, it appears that the lexicon is indeed flexible in the information it is able to store and utilize: though same-C confusions did not vanish, NN participants were able to utilize vowel information more heavily, as indexed by same-vowel errors, rather than showing an across-the-board increase in error rates over CC participants. Importantly, this suggests that word similarity undergoes a change depending on the listening conditions. Creel et al. [5] found related results with a different type of noise: segmentation difficulty. Confusions of words with an initial mismatched segment were more numerous as the segmentation difficulty increased. Second, noise affects both lexical storage and lexical access. Noise only at test and noise only at exposure both increased rates of same-V confusions over the CC condition. Third, the lexicon is specific. Error rates were greatest when exposure and test conditions differed (CN and NC conditions).

5. Conclusions

We have demonstrated both the robustness of the consonant in lexical storage and identification, and the flexibility of the lexicon in taking advantage of information present in the input. Most significantly, lexical similarity was more heavily weighted towards vowels when participants learned novel words in a background of white noise This work has implications for theories of lexical learning and lexical access. Noise level appears to affect both the nature of the lexical entry and the nature of lexical confusions. For theories of spoken word recognition, this research suggests that the acoustic environment affects the degree of perceived similarity among words—a factor that has not been taken into account in previous models of the lexicon [1,2,3,14]. Future research will examine more specifically the effects of different types of noise on lexical similarity and confusion patterns.

6. Acknowledgments

This work was supported by NIH grant DC 005071 to MKT and RNA, a Packard grant 2001-17783 to RNA, and an NSF graduate research fellowship to SCC.

7. References

[1] Gaskell, G., & Marslen-Wilson, W. D. "Phonological variation and inference in lexical access", *J Exp Psychol Human*, 22:144-158, 1996.

[2] McClelland, J. L. & Elman, J. L. "The TRACE model of speech perception", *Cog Psych*, 18:1-86, 1986.

[3] Norris, D. "Shortlist: A connectionist model of continuous speech recognition", *Cognition*, 52:189–234, 1994.

[4] Salverda, A. P., Dahan, D., & McQueen, J. M. "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension", *Cognition*, 90:51-89, 2003.

[5] Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. "Consequences of lexical stress on learning an artificial lexicon", under revision.

[6] Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. "Time course of frequency effects in spoken-word recognition: evidence from eye movements", *Cognitive Psychol*, 42:317-367, 2001.

[7] Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. "The time course of spoken word learning and recognition: studies with artificial lexicons", *J Exp Psychol Gen*, 132:202-227, 2003.

[8] McMurray, B., Tanenhaus, M. K., & Aslin, R. N. "Gradient effects of within-category phonetic variation on lexical access", *Cognition*, 86:B33-B42, 2002.
[9] Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. "The roles of segment type and order information in acquiring a miniature lexicon", under review.
[10] Cutler, A., Weber, A., Smits, R., & Cooper, N. "Patterns of English phoneme confusions by native and non-native listeners", *J Acoust Soc Am*, 116:3668-3678, 2004.
[11] Dorman, M. F., Kewley-Port, D., Brady, S., & Turvey, M. T. "Vowel recognition: inferences from studies of forward and backward masking", *Q J Exp Psychol*, 29:483-497, 1977.
[12] Goldinger, S. D. Echoes of echoes? An episodic theory of lexical access", *Psych Rev*, 105:251-279, 1996.
[13] Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. "PsyScope: A new graphic interactive environment for designing psychology experiments", *Behav Res Meth Instr*, 25:257-271, 1993.
[14] Luce, P. A., & Pisoni, D. B. "Recognizing spoken words: the neighborhood activation model", *Ear Hearing*, 19:1-36, 1998.