THE INFLUENCE OF NOISE ON PHONOLOGICAL COMPETITION DURING SPOKEN WORD RECOGNITION

Susanne Brouwer & Ann R. Bradlow

Northwestern University, USA
s-brouwer@northwestern.edu; abradlow@northwestern.edu

ABSTRACT

Listeners' interactions often take place in auditorily challenging conditions. We examined how noise affects phonological competition during spoken word recognition. In a visual-world experiment, which allows us to examine the time-course of recognition, English participants listened to target words in quiet and in noise while they saw four pictures on the screen: a target (e.g. *candle*), an onset overlap competitor (e.g. *candy*), an offset overlap competitor (e.g. *sandal*), and a distractor. The results showed that, while all competitors were relatively quickly suppressed in quiet listening conditions, listeners experienced persistent competition in noise from the offset competitor but not from the onset competitor. This suggests that listeners' phonological competitor activation persists for longer in noise than in quiet and that listeners are able to deactivate some unwanted competition when listening to speech in noise. The well-attested competition pattern in quiet was not replicated. Possible methodological explanations for this result are discussed.

Keywords: spoken word recognition, speech in noise, eye-tracking

1. INTRODUCTION

Research in the area of spoken word recognition has mainly used carefully pronounced speech presented in quiet conditions to show how fast and efficient listeners analyze the continuous speech signal over time [see 7]. This type of research has also shown that the recognition of spoken words involves continuous activation of multiple lexical candidates [5]. Candidates that sound similar to the speech signal are activated and compete with each other until they mismatch with the input. For example, the word *candle* will activate *candy* but not *lemon*.

The visual-world paradigm has been found to be a useful technique to investigate such phonological competition effects [8]. In this method, listeners' eye movements are measured as they listen to speech and see pictures of objects on a screen. The timing and proportion of fixations to pictures of objects reveal which lexical candidates the listener is entertaining as speech unfolds over time. In previous work participants' eye movements were tracked to four pictures on a screen (e.g., a candle, a candy, a sandal, and a lemon), while they had to follow spoken instructions such as "Pick up the candle" [1]. This study showed that in the period before listeners settled unequivocally on the target word, they fixated more on pictures that matched the target signal (*candy*, *sandal*) than those that were phonologically unrelated to the target signal (*lemon*). Importantly, participants looked more often to competitors matching at word onset (cohort competitor: *candy*) than competitors matching at word offset (rhyme competitor: *sandal*).

The present study examines to what extent these results can be generalized to adverse listening situations. To our knowledge only two studies examined the effects of noise on competition effects in a visual-world paradigm [3, 6]. Both studies used a variant of [1] in which the onset and offset overlap competitors were independently displayed on the screen with a target and two distractors. The first study on competition effects in noise investigated age-related differences in spoken word processing [3]. Younger and older adults listened to blocked presentations of target sentences in quiet or in noise. We focus here on the eye movement results for the younger adults as they are comparable to our population. Their data showed that, at later time points, noise had a larger influence on the discrimination of targets from onset overlap competitors compared to offset overlap competitors. For offset overlap trials this effect occurred earlier and lasted only for a short amount of time.

The other eye-tracking study used a between subjects design to compare spoken word recognition in quiet with a condition in which the onset phonemes of words in target sentences were replaced with radio-signal noises [6]. Note that the noises never occurred in the target words themselves, which may underestimate what happens when the target itself is masked by noise. The results in the quiet condition replicated previous work [1]. In the noise condition, participants still fixated on onset competitors more than on offset competitors but the early onset effect was reduced and the late offset effect was stronger and occurred earlier. These results suggest that the dynamics of spoken word recognition are modulated by noise. When onset information is less reliable, listeners seem to adjust their interpretation of the acoustic signal.

The main goal of the current study was to replicate and extend the previous work with two significant methodological modifications. First, we displayed both types of competitors on the screen in one trial to examine competition effects more directly. That is, the separate presentation of onset and offset competitors in the previous studies may have constrained the full picture of how candidate word competition unfolds over time. In combination with the auditory uncertainty introduced by noise, the uncertainty in the response set introduced by the presence of two competitors may affect processing dramatically differently from the case with just one competitor present.

Secondly, we introduced even more uncertainty by presenting a mixed design (i.e. mix of quiet and noise trials), which prevents participants from adjusting to the noise. Different competitor activations in blocked and mixed design have also been found in previous work [4]. In that study, a preference for onset competitors was found when all target words were fully pronounced, but not when carefully pronounced targets were intermixed with reduced targets. This suggests that listeners penalize acoustic mismatches less strongly when the listening context as a whole is non-optimal by including reduced speech. In the current study we investigate whether this broad context effect may also hold when background noise is added on some trials.

These methodological changes are important for a full understanding of speech recognition under adverse conditions. By increasing uncertainty in the response set and by introducing noise for some trials, we created an overall testing paradigm that more closely resembles real-world listening situations. We expected that these manipulations would influence the competition pattern of previous work in the quiet [1] and in the noise condition [3] such that only the strongest competitions effects would be observed, namely onset competition.

2. EXPERIMENT 1

2.1. Method

2.1.1. Participants

We tested 26 monolingual American-English listeners (9 males, 17 females) with normal hearing and with normal or corrected-to-normal vision.

2.1.2. Material

Sixteen disyllabic nouns referring to picturable objects were selected as targets. Each target was paired with two competitors. The onset overlap competitor matched with the initial sounds of the target (e.g. ‘candy’ for the target ‘candle’) and the offset overlap competitor matched with the final sounds of the target (e.g. ‘sandal’ for the target ‘candle’). The target and competitor overlapped minimally with one syllable. On a given trial, the phonologically related objects were displayed with a phonologically unrelated distractor (e.g. lemon). The quadruplets were matched on CELEX estimates of frequency of occurrence.

Each display was presented along with a target word which was recorded by a female native speaker of American-English, sampling at 22050 Hz. The recordings were equalized to the same rms level. The level of the target words was fixed at 65 dB SPL. For the noise condition, targets were mixed in PRAAT© with speech-shaped noise. The noise was played at 67 dB (SNR level of -2 dB).

An additional 16 quadruplets were selected for filler trials consisting of three phonologically related and one unrelated word. For these trials, the target could be any of the words, preventing participants from developing strategies regarding the mention of items sharing phonological attributes. Finally, 8 similarly-constructed practice quadruplets were selected.

2.1.3. Procedure, design, and analysis

Prior to the experiment, participants were shown pictures of the stimuli they were to see in the experiment and asked in a 2AFC task which of two printed words represented the picture. Results showed that participants made no errors.

Participants’ eye-movements were monitored at a sampling rate of 1 kHz with an SR Research

EyeLink1000 eye-tracker (used in the tower-mounted version). The presentation of the auditory and visual stimuli was controlled with SR Research program Experiment builder. The auditory stimuli were presented over headphones. After a calibration procedure, participants received written instructions on the screen. They were asked to click on the picture in the visual display representing the word they heard. This word could be presented in quiet or in the presence of noise.

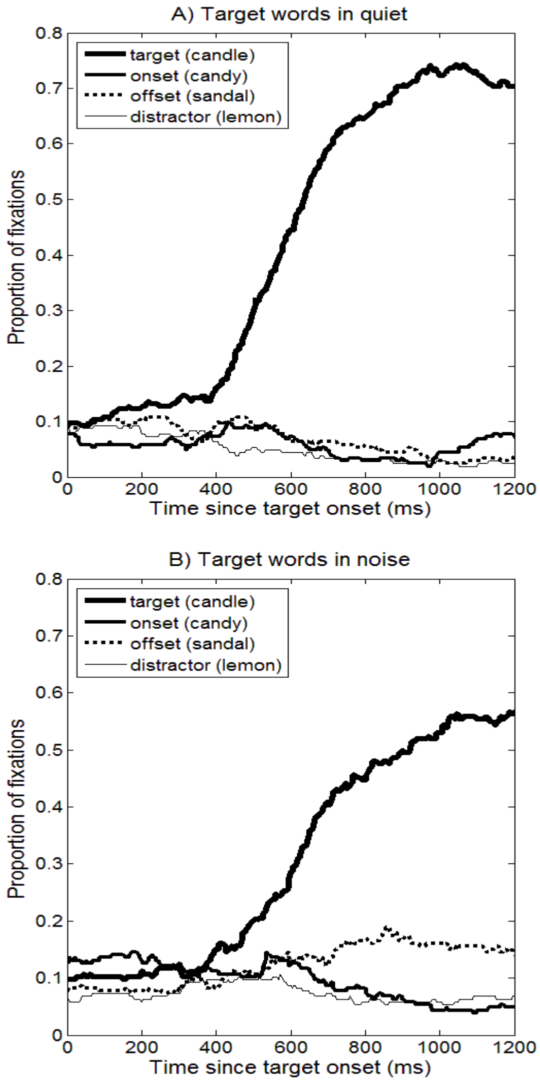
The location of the pictures was randomized over the four quadrants. On each trial, the four pictures were first displayed. After 1000 ms, the auditory stimulus was presented. After a practice session of 8 trials, participants were presented with a total of 32 experimental and filler trials. Quiet and noise trials were randomly mixed. Two different item lists were created. Both lists contained half of the targets in quiet and half of the targets in noise. Each participant received one list. The trials in each list were randomized. When participants clicked with the mouse on a word, they initiated the next trial. After every ten trials, a central fixation cross appeared centered on the screen so that the experimenter could correct drifts in the calibration of the eye-tracker. The experimental session took circa 10 minutes.

The eye-tracking data were analyzed using linear mixed effects models [2]. The mean fixations were transformed into empirical logits to create three linearly independent measures: 1) looks to the target, to investigate the ease of recognition; 2) looks to both competitors vs. the distractor, to assess the existence and strength of overall competition effects; and 3) looks to the onset overlap competitor vs. offset overlap competitor, to test for the specificity of the competition effects. We tested whether these measures were influenced by Noise Type (Quiet vs. Noise) as fixed effect and with participants and items as random effects. Noise Type was coded as a numeric contrast (-0.5 and +0.5). The quiet condition was coded as -0.5 and the noise condition as +0.5. A negative regression weight (beta) implies more fixations in the quiet than in the noise condition. We computed the mean fixations in an early (400-800 ms) and in a late time window (800-1200 ms).

2.2. Results

Figure 1 shows the proportion of fixations over time from 0 to 1200 ms after target word onset for A) targets in quiet and B) targets in noise. The target analysis showed a main effect of Noise Type (400-800: $\beta_{NoiseType}$ = -1.63, $p_{MCMC} < 0.001$; 800-1200: $\beta_{NoiseType}$ = -1.87, $p_{MCMC} < 0.001$). The negative betas indicate that listeners looked more often at targets in quiet than in noise.

Figure 1: Fixation proportions over time from target onset for target words in quiet (A) and in noise (B).



Further, we analyzed whether listeners looked more at the two types of competitors than the distractor. This analysis showed an effect of overall competition (400-800:$\beta_{Intercept}$ =0.37, $p_{MCMC} < 0.05$; 800-1200:$\beta_{Intercept}$ =0.51, $p_{MCMC} < 0.05$), independent of Noise Type (all $p_{MCMC} > 0.1$).

Finally, a comparison between looks to the two competitors showed significant differences in the late time window only. Looks to the competitors differed from each other ($\beta_{Intercept}$ = -0.51, p_{MCMC} = 0.05) and this effect was modulated by Noise Type

($\beta_{NoiseType}$ = -1.21, p_{MCMC} < 0.001). The negative beta indicates that listeners looked more often at the offset overlap competitor in noise than in quiet.

3. DISCUSSION

In an eye-tracking experiment we extended previous work on phonological competition in quiet and in noise. We found that listeners recognize targets more slowly when presented with speech in noise than in quiet, revealing an overall processing cost in noise. Moreover, persistent offset competition was found indicating that listeners activate lexical candidates that share offset overlap with the target until a late moment in time. However, onset competition was not found, indicating that listeners are able to deactivate some unwanted competition. These findings are inconsistent with previous work [3], indicating that the simultaneous visual presentation of two competitors and a mixed design influence the phonological competition pattern.

The well-established competition pattern for targets in quiet, that is, a preference for onset competitors over offset competitors early on in the decision process was not replicated. There are three possible explanations for this conflicting finding. First, it could be that listeners interpret the acoustic signal with more flexibility because they hear a mix of auditorily clear speech (quiet condition) and degraded speech (noise condition). As found in previous work with reduced speech [4], listeners may adjust to the most difficult condition, in this case the noise condition, resulting in a decrease of the criterion by which target-to-competitor mismatches are penalized. To test this possibility, we are conducting a follow-up experiment with blocked conditions (first quiet, then noise).

A second explanation for the discrepancy between our results and the classic competition pattern is our choice of visual display preview time. In our study, participants had 1000 ms to preview the visual display with the four pictures before onset of the auditory stimulus. Although this preview time is common in eye-tracking studies [1], other studies use a shorter preview time [6]. Our longer preview time could have induced certain strategies. For example, it could be the case that listeners first quickly scan the screen with the four pictures. When the target word subsequently unfolds, they are immediately able to make the match with the correct picture. This could have weakened the competition effects considerably especially in quiet. In a follow-up experiment we test this possibility by modifying the preview time.

A final explanation is that the auditory targets in our study were presented in isolation, whereas previous work presented targets following a carrier phrase [1, 3]. We presented isolated targets because, as a next step, we plan to substitute carefully constructed background speech noise for broadband noise in an effort to test lexical competition patterns within the auditory modality rather than just in the visual display. The lack of a carrier phrase could affect competitor activation by removing the benefit of predictability.

In conclusion, the current results showed that the dynamics of spoken word recognition are influenced by extrinsic noise. Specifically, lexical candidates that share offset overlap with the target are highly activated at later moments in time when noise is present. The fact that onset competition was essentially absent from the quiet and the noise condition suggests that even this type of competition is somewhat fragile under conditions of uncertainty. Clearly, examining phonological competition under different listening conditions offers a promising avenue for future research.

4. REFERENCES

[1] 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. *J. Mem. Lang.* 38, 419-439.

[2] Baayen, R.H., Davidson, D.J., Bates, D.M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390-412.

[3] Ben-David, B.M., Chambers, C.G., Daneman, M., Pichora-Fuller, K.M., Reingold, E.M., Schneider, B. 2010. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *JSLHR* 54, 243-262.

[4] Brouwer, S., Mitterer, H., Huettig, F. In press. Speech reductions change the dynamics of competition during spoken word recognition. *Proc. Lang. Cogn.*

[5] Goldinger, S.D., Luce, P.A., Pisoni, D. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *J. Mem. Lang.* 28(5), 501-518.

[6] Huettig, F., McQueen, J.M. 2009. AM Radio Noise Changes the Dynamics of Spoken Word Recognition. Paper presented at the *AMLaP* conference in Barcelona, Spain.

[7] McQueen, J.M. 2007. Eight questions about spoken-word recognition. In Gaskell, M. (ed.), *The Oxford handbook of psycholinguistics.* Oxford: OUP, 37-53.

[8] 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.