The role of phoneme order and phonetic detail in spoken word recognition

Joseph C. Toscano (jtoscano@illinois.edu)
Beckman Institute, University of Illinois at Urbana-Champaign

Nathaniel D. Anderson (nandrsn3@illinois.edu)
Dept. of Psychology and Beckman Institute, University of Illinois at Urbana-Champaign

Bob McMurray (bob-mcmurray@uiowa.edu)
Dept. of Psychology and Dept. of Communication Sciences & Disorders, University of Iowa

A basic challenge in understanding spoken word recognition is that speech unfolds over time. This has led to a great deal of work in psycholinguistics on how listeners deal with temporary ambiguity (Allopenna, Magnuson, & Tanenhaus, 1998; Luce & Pisoni, 1998; Marslen-Wilson, 1987), demonstrating that during early time points in a word (when its identity is still ambiguous), listeners consider multiple lexical candidates that compete for recognition. Thus, after hearing /tæ/, listeners may consider *tack*, *tap*, and *taxi* (*cohorts*, words with the same onset) as possible completions. One consequence of this is that all models of spoken word recognition assume that the order of phonemes in a word is essential to the recognition process. This suggests that words like cohorts compete for activation, but words like *phonemic anadromes* (words with the same phonemes but opposite temporal order; e.g., *cat* vs. *tack*) do not.

Is this assumption necessary? An alternative is that words may be defined by temporal order only loosely or not at all. Because phonemes vary acoustically with word position (e.g., the /k/ in *cat* is different from the /k/ in *tack*), listeners may be able to distinguish words without needing to track the order of phonemes in them. This idea is supported by work demonstrating that listeners are sensitive to fine-grained acoustic detail during word recognition (McMurray, Tanenhaus, & Aslin, 2002; Ranbom & Connine, 2007; Salverda, Dahan, & McQueen, 2003). In addition, words sharing the same phonemes will have some acoustic similarities, which, in contrast to the predictions of existing models, could lead listeners to consider anadromes.

We examined whether this was the case using a visual world eye-tracking experiment in which we measured participants' eye-movements to objects on a computer screen while they selected the object that corresponded to a spoken word. The display contained four objects comprising an item-set: a target (e.g., *tack*), its anadrome (*cat*), a cohort competitor (*tap*), and a phonetically-unrelated item (*mill*). Listeners were more likely to look at anadromes than unrelated objects (Figure 1), indicating that they considered them even though the phoneme order did not match the auditory stimulus. We also ruled out the possibility that this effect was driven by temporal overlap with the vowel (e.g., *tack* and *cat* both share the vowel /æ/ in the correct position) by examining fixations to competitors that shared a vowel and another phoneme (e.g., *tap* when the stimulus was *cat*). We found that fixations to these objects were not significantly different from fixations to the unrelated objects. In addition, later in the trial, there were more fixations to anadromes than to vowel-overlap objects. Finally, when we controlled for differences in fixations due to the visual properties of the objects, we found significantly more fixations to anadromes than to vowel-overlap or unrelated objects (Figure 2).

These results show that listeners consider anadromes during word recognition and that lexical activation is not completely determined by phoneme order. This presents a challenge to all existing models that make this assumption and suggests that we should reconsider how temporal order is implemented in models of spoken word recognition. It also suggests that a way to move forward may be to incorporate additional fine-grained acoustic detail into models.
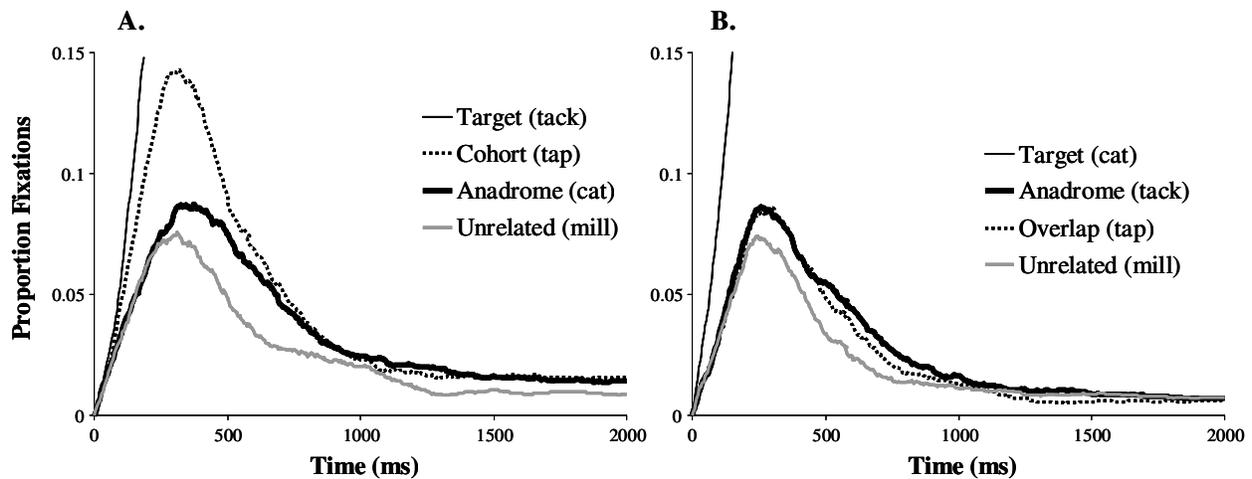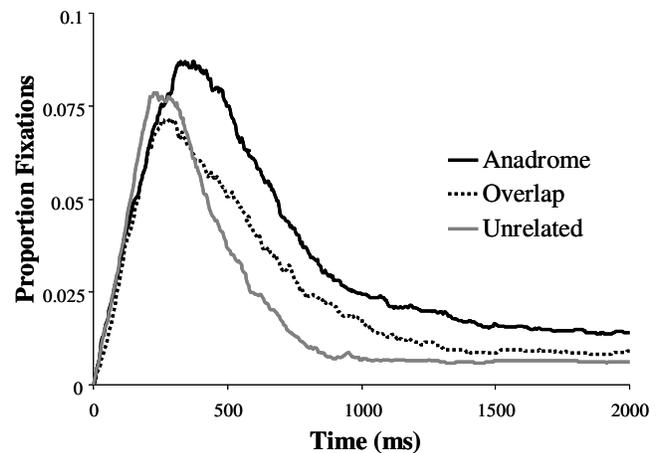
Figure 1 (above). Proportion of fixations each type of object over the course of the trial on trials containing (A) cohort and anadrome competitors and (B) anadrome and overlap competitors.

Figure 2 (right). Proportion of fixations to a given object when it served as a different type of competitor.

References

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

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33-B42.

Ranbom, L., & Connine, C. M. (2007). Lexical representation of phonological variation. *Journal of Memory and Language, 57*, 273-298.

Salverda, A. P., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51-89.