Reconsidering the role of temporal order in spoken word recognition

Joseph C. Toscano · Nathaniel D. Anderson ·
Bob McMurray

Published online: 28 February 2013
© Psychonomic Society, Inc. 2013

Abstract Models of spoken word recognition assume that words are represented as sequences of phonemes. We evaluated this assumption by examining *phonemic anadromes*, words that share the same phonemes but differ in their order (e.g., *sub* and *bus*). Using the visual-world paradigm, we found that listeners show more fixations to anadromes (e.g., *sub* when *bus* is the target) than to unrelated words (*well*) and to words that share the same vowel but not the same set of phonemes (*sun*). This contrasts with the predictions of existing models and suggests that words are not defined as strict sequences of phonemes.

Keywords Spoken word recognition · Speech perception · Temporal order · Eyetracking · Visual-world paradigm

A significant problem in understanding spoken word recognition is that speech unfolds over time. This leads to *temporary ambiguity*: Information early in the signal is often

insufficient to identify the intended word, since it is consistent with many different words (Marslen-Wilson, 1987). For example, when hearing *tack*, after only /tæ/ is heard, many completions are possible (e.g., *tack*, *tap*, *taxi*). A related issue involves *temporal order*: The order of elements in a word seems to be important for distinguishing them. For example, *sub* and *bus* consist of the same phonemes but can be distinguished because those phonemes occur in different sequences. There has been little research on the effects of temporal order, despite the assumption that it is fundamental to lexical representations, and the fact that it is implemented quite differently in different models (Gaskell & Marslen-Wilson, 1997; Grossberg, 2003; McClelland & Elman, 1986; Norris & McQueen, 2008). In the present study, we examined its role in spoken word recognition by asking whether phonemes in incorrect positions contribute to lexical access.

Previous studies on temporary ambiguity are relevant to this problem. These studies have led to a consensus that listeners access potential lexical candidates from the earliest moments of a word (Alloppenna, Magnuson, & Tanenhaus, 1998; Marslen-Wilson & Zwitserlood, 1989); they consider multiple words in parallel (Luce & Pisoni, 1998; Marslen-Wilson, 1987); they update the words under consideration as subsequent information arrives (Dahan & Gaskell, 2007; Frauenfelder, Scholten, & Content, 2001); and words compete with each other for recognition (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998).

This work also makes the strong assumption that words are defined as phoneme *sequences*.¹ This seems intuitive: If phoneme order distinguishes words, it makes sense for the system to represent them in this way. The large body of research showing strong competition from onset-matching

Electronic supplementary material The online version of this article (doi:10.3758/s13423-013-0417-0) contains supplementary material, which is available to authorized users.

J. C. Toscano (✉) · N. D. Anderson
Beckman Institute for Advanced Science and Technology,
University of Illinois at Urbana-Champaign, 405 N. Mathews Ave,
Urbana, IL 61801, USA
e-mail: jtoscano@illinois.edu

N. D. Anderson
Department of Psychology, University of Illinois
at Urbana-Champaign, Champaign, IL, USA
e-mail: nandrsn3@illinois.edu

B. McMurray
Department of Psychology and Department of Communication
Sciences and Disorders, University of Iowa, Iowa City, IA, USA
e-mail: bob-mcmurray@uiowa.edu

¹ We use *phonemes* here only as a convenience to describe the input; we are not implying any representational status for them in the recognition process.

competitors (cohorts) provided some of the earliest evidence for this assumption. Since cohorts are more active than other competitors during early portions of the signal, this result appears to support slot-based representations in which early parts of the signal are matched to early parts of words in the lexicon. Indeed, the COHORT model (Marslen-Wilson, 1987) suggested that recognition was all-or-none: Words that did not match at onset did not compete for activation. However, empirical work showing competition from onset-mismatching words has challenged this assumption (Allopenna et al., 1998; Connine, Blasko, & Titone, 1993). As a consequence, current models allow activation to depend on the degree of match in each phoneme position, but continue to use a slot-based scheme (Gaskell & Marslen-Wilson, 1997; Luce, Goldinger, Auer, & Vitevitch, 2000; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris & McQueen, 2008). In NAM and Shortlist B, for example, graded activation is implemented by using phoneme confusion data to partially activate words that are phonologically similar within each slot. In TRACE, this is accomplished through the use of graded features for each phoneme. However, while current models allow for this graded activation, they all incorporate the idea that the input is matched to lexical items slot-by-slot.

In the present study, we challenged this assumption, asking whether words are even represented as sequences. This research was loosely inspired by work on *visual* word recognition suggesting that letter order may be only coarsely coded (Chambers, 1979; Grainger & Whitney, 2004). Nonwords with transposed letters (JUGDE) prime their original words (JUDGE), while the same-sized mismatch without transposition (JULPE) does not (Perea & Lupker, 2003); this also extends to nonadjacent transpositions (e.g., CANISO/CASINO: Perea & Lupker, 2004). This implies that letters in incorrect positions may still activate the correct target word. This does not mean that readers completely ignore letter order, and there are limitations to these effects (Guerrara & Forster, 2008; Hannagan, Dupoux, & Christophe, 2011). However, the effect suggests that printed words are not represented as strict letter sequences. While spoken and written words differ in both temporal demands and the nature of the input (the forms of letters are the same in all positions, while the forms of phonemes are not), this raises the possibility that order need not be fully represented.

Abandoning a slot-based approach for spoken words may appear to create problems (e.g., distinguishing *sub* from *bus*). However, given that phonemes vary acoustically with word position and that listeners are sensitive to fine-grained differences (McMurray, Tanenhaus, & Aslin, 2002), people may not need to represent words in a slot-based format: *Sub* and *bus* may be distinguished because word-initial /b/ is acoustically different from word-final /b/. Under such a

system, listeners might show strong cohort effects, since when they hear a word-initial /b/, its acoustic properties map more strongly onto words with the word-initial allophone of /b/, and less well onto words with the word-final allophone. Critically, these effects would not arise from slot-based representations.

We tested this possibility by examining *phonemic anadromes*: words like *sub* and *bus* that contain the same phonemes in the opposite order. Since many models tolerate some mismatch at onset (Gaskell & Marslen-Wilson, 1997; Luce et al., 2000; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris & McQueen, 2008), they might predict activation for anadromes over unrelated words that share no phonological features (on the basis of the shared vowel). However, they predict that this should be the same as activation for nonanadromes that have a similar degree of mismatch (e.g., *sun* would compete with *bus* just as well as *sub* does). None predict competition from anadromes due to phonemes in incorrect positions (e.g., the /b/ in *sub* leading to activation for *bus*). We have confirmed that this is the case for TRACE (see the supplemental materials, section S1). Thus, existing models suggest that *sub* might compete weakly with *bus* because of the similarity of the words *within* each slot, since the matching vowels (in the second slot) could drive some activation (though it has never been empirically demonstrated that a single word-medial phoneme can drive competition). Given this, we must also compare anadromes (*sub* when *bus* is the target) to words having the same onset and vowel (*sun*).

Subphonemic overlap among the consonants could also play a role. While *sub* and *bus* have little overlap, consider *tack* and *cat*. In TRACE, the features for /t/ are similar to those for /k/. In NAM and Shortlist B, the /t/ would be partially confusable with /k/. Therefore, any word with some confusability in each slot could compete. Thus, *tap* could compete with *cat*, and if /p/ and /k/ are equally distant from /t/, *tap* and *tack* could be equally active. Critically, this activation still derives from slot-based representations; phonemes or features in the wrong slot do not drive the effect. Thus, to establish anadrome activation, we would also need to look at word pairs containing initial consonants with minimal phonological overlap.

We used the visual-world paradigm (Allopenna et al., 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to ask whether anadromes are activated during lexical access, and if so, whether part of this activation can be attributed to phonemes whose positions in the input do not match the corresponding positions in the competitor word. Participants heard a target word in the presence of four pictures representing the target, possible lexical competitors, and unrelated words. Each item set contained a base word (e.g., *sub*), its anadrome (*bus*), a cohort (*sun*), and an unrelated word (*well*). Critically, on some trials, anadromes

could be compared to words that overlapped only in their vowel (e.g., when *bus* was the target, differences in looks to *sub* and *sun*), allowing a test of effects of the vowel and of partially matching initial consonants. While ideally the word list would include only anadromes with minimal overlap among the bracketing consonants, sufficient picturable pairs were not available in English to construct a large list. Thus, our list consisted of words with similar initial consonants (e.g., *tack/cat*), which were useful for establishing activation for anadromes, and words that were minimally overlapping (*sub/bus*), which could confirm that effects were driven by phonemes in different positions.

Method

Participants A group of 30 University of Iowa undergraduates participated. The participants were native English speakers, reported normal or corrected-to-normal vision, provided informed consent, and received course credit or monetary compensation.

Design Participants heard spoken words and used a computer mouse to select a corresponding picture. The stimuli consisted of 16 sets of four words (see the [Appendix](#)). Each set contained a base word (e.g., *sub*), its anadrome (*bus*), a cohort (*sun*), and an unrelated word (*well*).

Across trials, each word in an item set occurred an equal number of times. This created four trial types. On cohort/anadrome trials (in which the base word—e.g., *sub*—was the target), cohort and anadrome competitors were present with the target and an unrelated word. On anadrome/overlap trials (e.g., *bus* was the target), an anadrome (*sub*, the former base word) and an unrelated word (*well*) were present; the former cohort (*sun*) was termed an *overlap*, since it shared only an overlapping vowel and a consonant in the wrong position. Cohort/overlap trials (e.g., *sun* was the target) contained a cohort (*sub*), an overlap (*bus*), and an unrelated word (*well*). Finally, on unrelated trials (e.g., *well* was the target), all three competitors were phonologically unrelated. Table 1 shows an example item set and the roles of each word across trial types. Each word was repeated nine times, yielding 576 trials. Participants completed the experiment in an hour-long session.

Table 1 Example item set, by trial types

Trial Type	<i>sub</i>	<i>sun</i>	<i>bus</i>	<i>well</i>
Cohort/anadrome	Target	Cohort	Anadrome	Unrelated
Anadrome/overlap	Anadrome	Overlap	Target	Unrelated
Cohort/overlap	Cohort	Target	Overlap	Unrelated
Unrelated	Unrelated	Unrelated	Unrelated	Target

Stimuli Auditory stimuli were recorded by a female talker in a sound-attenuated room. Recordings were made at 44.1 kHz on a Kay Elemetrics Computerized Speech Lab 4300B. Several tokens of each word were recorded, and the best exemplar was selected. The mean word duration was 457 ms ($SD = 74$ ms).

The visual stimuli were color drawings prepared using a standard procedure in the McMurray lab. For each word, a set of candidate clip-art images were obtained. These were viewed by focus groups of lab personnel to select the most prototypical image and to guide subsequent editing to make pictures representative and uniform (McMurray, Samelson, Lee, & Tomblin, 2010).

Procedure An SR Research EyeLink II eyetracker was calibrated using the standard nine-point procedure. On each trial, participants saw four 200×200 pixel pictures in each corner of a 19-in. CRT monitor. At the center, a small blue dot was displayed, which turned red after 750 ms. The participant then clicked on the dot and heard the auditory stimulus 100 ms later (via Sennheiser HD555 headphones). This ensured that both the mouse cursor and the participant's gaze were centered when the auditory stimulus began. The trial ended when the participant clicked on the corresponding referent.

Every 45 trials, drift correction was performed, and participants were given the opportunity to take a break. Stimulus presentation and data collection were handled by the SR Research Experiment Builder package and by the EyeLink control software.

Eye-movement analysis As in prior experiments (McMurray et al., 2010; McMurray et al., 2002), the eye-movement record was parsed into saccades and fixations by the EyeLink software. Adjacent saccades and fixations were combined into a “look” extending from saccade onset to fixation offset. The boundaries around the images were extended by 100 pixels. This maintained substantial space between ports (horizontal, 630 pixels; vertical, 374 pixels) while compensating for noise.

We used a standard area-under-the-curve approach for the data analysis (Allopenna et al., 1998; McMurray et al., 2002), with the time window starting 200 ms after stimulus onset (since it takes ~200 ms to plan an eye movement) and ending at 1,417 ms (the average response time [RT] plus 200 ms).² Proportions of looks were analyzed with linear mixed-effects (LME) models using the lme4 package in R

² For trials longer than 1,417 ms, looks were truncated. For trials shorter than 1,417 ms, the final fixation position was used to fill the remaining time period, assuming that listeners had settled on an interpretation when they clicked on the referent. This approach has been used in several previous studies (McMurray et al., 2010; McMurray et al., 2002).

(Bates & Sarkar, 2011). The dependent variable was proportion of looks, transformed with the empirical logit for use with linear models. All models used object type as the only fixed effect, which was binary and coded as $-.5/+1.5$.

In considering the random-effects structures of the models, we found that adding by-subject slopes did not significantly improve model fit, while by-item random slopes did. Models with only random intercepts for items showed clearer effects; models that were more sensitive to item-level variation using random slopes did not uniformly show effects. This suggests that our effects generalized robustly across subjects, but not necessarily across items. This was not unexpected, since our design does not provide sufficient power to detect small by-item effects, and the items shared a number of sources of variability. Thus, we report here results both from models with only random intercepts and from models with random slopes (see the supplemental materials, section S2, for complete details of all statistical models). To be clear, both models included random effects for subjects and items. Significance was evaluated using the chi-square goodness-of-fit test comparing models with and without object type as a fixed effect.

Results

Participants performed well, averaging 99.3 % correct ($SD = 1.1$ %). We eliminated the few trials on which they selected the incorrect referent ($M = 3.9$ trials/participant, $SD = 6.5$). The mean RT for correct trials was 1,217 ms ($SD = 642$ ms).

To determine whether anadromes competed for recognition, we first examined the proportions of looks to the anadrome and unrelated objects on cohort/anadrome trials. Figure 1, left panel, shows the time course of looks to each competitor on these trials. The earliest fixations were directed to the target and cohort, since at that time, the stimulus was consistent with both. Fixations to the anadrome were initiated later and were greater than those to the unrelated object. This was confirmed statistically in an LME model with random intercepts and object type (anadrome vs. unrelated) as the only fixed effect ($b = 0.025$, $SE = 0.006$, $p < .001$), although this was not significant in the random-slope model ($b = 0.025$, $SE = 0.017$, $p = .131$).

Since words that mismatch at onset can also be activated (Allopenna et al., 1998), the anadrome effect could be driven by the overlapping vowel and/or by shared features at onset (e.g., the /k/ in *cat* shares features with the /t/ in *tack*, since both are voiceless stops). Such a model also predicts activation for *tap* (an overlap competitor) when *cat* is the target. We examined this with a second LME model, which compared looks to overlap and unrelated objects on cohort/overlap trials (Fig. 1, right). The proportions of looks to the two objects were not significantly

different in either the random-intercept model ($b = 0.001$, $SE = 0.006$, $p = .847$) or the random-slope model ($b = 0.001$, $SE = 0.006$, $p = .847$). This suggests that overlaps were not activated more than unrelated items.

We next asked whether anadromes also received more looks than overlaps. Here, we considered two approaches to analyzing this difference. One follows the analyses above, which examined looks within a single trial type, holding the auditory stimulus constant and examining fixations to different visual objects in the display. An alternative would be to compare looks to the same visual object when it served as a different type of competitor. For example, we could compare looks to *bus* when it served as an anadrome on cohort/anadrome trials (with *sub* as the target) with looks to *bus* as an overlap on cohort/overlap trials (with *sun* as the target). This controlled for differences in the visual stimulus that might drive differences in eye movements, gave us twice as many trials from which to draw our data, and might make it easier to detect small competitor effects, since it used trials with a cohort in the display (Dahan et al., 2001).³

First, we used the within-trial-type approach to look at differences between anadromes and overlaps on anadrome/overlap trials (Fig. 2, left). We found a marginal effect of object type, with more looks to anadromes than to overlaps in the random-intercept model ($b = 0.011$, $SE = 0.006$, $p = .092$), which appeared to be restricted to a few time ranges (300–500 and 800–1,200 ms); this effect was not significant in the random-slope model ($b = 0.011$, $SE = 0.016$, $p = .523$). One reason that we might not see clear anadrome–overlap differences here could be that, for many item sets, the competitor onsets shared multiple features with the target, which could heighten overlap activation. If *cat* was the target, the initial phonemes of *tack* and *tap* (anadrome and overlap) differed from *cat* in just one feature. For these item sets, we might expect that overlaps would be activated more by their shared phonological features with the target.

Therefore, a stronger test of our hypothesis would be to examine differences in anadrome and overlap looks for pairs whose consonants shared no features (e.g., for *bus* [target], *sub* [anadrome], and *sun* [overlap], the competitors' initial phoneme differed from the target's in voicing, place, and manner). For these item sets, activation for the overlap and anadrome would not be enhanced by shared features at onset. If we were to observe more activation for anadromes than for overlaps in these item sets, it would suggest that competition was driven by the presence of phonemes that matched the target but were in the wrong position.

³ For the anadrome/unrelated and overlap/unrelated comparisons, this required us to compare trials with a cohort to those without one. Thus, we initially examined the within-trial-type comparisons for those effects.

Fig. 1 (Left) Time courses of looks to each object on cohort/anadrome trials. (Right) Time courses of looks on cohort/overlap trials. Dashed lines indicate the boundaries of the analysis time window

A fourth LME model examined this possibility by comparing anadrome and overlap looks on anadrome/overlap trials for the four item sets that differed in all three types of phonological features (Fig. 2, right). The random-intercept model showed an effect of object type, with more looks to anadromes ($b = 0.055$, $SE = 0.015$, $p < .001$); this effect was not significant in the random-slope model ($b = 0.055$, $SE = 0.040$, $p = .183$). This suggests that our effects cannot be due solely to graded activation of phonemes in slot-based representations of words; phonemes in the incorrect position play an additional role.

Finally, we examined four LME models corresponding to the comparisons above, but holding the visual stimulus constant and comparing across trial types (see Fig. 3). Each of the four random-intercept models showed an effect of object type: (1) anadrome > unrelated ($b = 0.044$, $SE = 0.006$, $p < .001$), (2) overlap > unrelated ($b = 0.016$, $SE = 0.005$, $p = .003$), (3) anadrome > overlap using all item sets ($b = 0.028$, $SE = 0.006$, $p < .001$), and (4) anadrome > overlap using the item sets with no feature overlap ($b = 0.046$, $SE = 0.012$, $p < .001$). Also, two of the random-slope models showed an effect: (1) anadrome > unrelated ($b = 0.044$, $SE = 0.018$, $p = .024$), and (2) anadrome > overlap using all item sets ($b = 0.028$, $SE = 0.013$, $p = .045$). Thus, these results show a clear advantage for anadromes over overlaps, and they also

reveal a small effect of overlap competitors that we did not observe in the within-trial-type analyses.

Discussion

We found robust evidence that anadromes are activated more than unrelated words during spoken word recognition. This is consistent with previous work showing activation for lexical competitors that mismatch at onset (e.g., rhymes), and adds to the set of competitors that listeners consider during word recognition.

There are two potential causes of this effect. One possibility is that words are represented in a slot-based way and that a matching phoneme in the second position (the vowel) and/or a mismatching, but similar, phoneme in the first position (the initial consonant) drove the effect. However, we only found mixed evidence that overlaps were activated at all, and we found much stronger evidence that anadromes received more fixations than overlaps when differences in the consonants were maximal. When overlap activation was not heightened by featural match between the overlap and the target, a clear anadrome effect could be seen.

This suggests that words are not represented in a slot-like format, and that phonemes in the wrong position may

Fig. 2 (Left) Time courses of looks to each object on anadrome/overlap trials. (Right) Time courses for the item sets with three feature differences between the onset of the target and competitors

Fig. 3 (Left) Time courses of looks to a given object when it served as a target (on anadrome/ overlap trials), anadrome (cohort/anadrome trials), and overlap (cohort/overlap trials). (Right) Time courses for item sets with three feature differences between target and competitor onsets

activate competitors (e.g., the /s/ in the first position of the input could activate *bus*). Thus, in contrast to existing models of word recognition, anadromes are activated as a consequence of phonemes in the incorrect position.

Given this, why don't listeners confuse words with their anadromes? One possibility is that fine-grained phonetic detail could serve as a proxy for temporal order. The acoustic forms of consonants vary with syllable positions: A syllable-initial /b/ contains rising formants and a short voice onset time, while a syllable-final /b/ contains falling formants, a closure, and a release burst. While both have similar gross spectral shapes (Blumstein & Stevens, 1979), they differ in many details. As a result, a fine-grained description of *sub* will differ from that for *bus*, even if phoneme order is ignored. This can explain why cohorts are more active than other competitors: A word-initial allophone of /b/ is a better match to *bus* and *bun* than to *sub*. We should note, however, that this is only a hypothesis at this point.

The fact that word recognition is sensitive to fine-grained detail (McLennan & Luce, 2005; McMurray et al., 2002), including allophonic detail (Ranbom & Connine, 2007), means that listeners could use subtle acoustic differences to distinguish anadromes. At the same time, the coarse spectral similarity would lead to some parallel activation. Such a model is quite different from most models of spoken word recognition, in that lexical representations have no inherent temporal order. Rather, words would be more like a bundle of fine-grained acoustic cues. This detailed acoustic information is preserved in exemplar models (Goldinger, 1998), although the memory traces for whole words are typically thought to include temporal order.

Although this differs from current models, it could be implemented in similar architectures using a model in which fine-grained cues were mapped directly to words (independently of order). A version of TRACE with a single time alignment and a richer input might show anadrome effects, as could architectures like normalized recurrence (Spivey, 2007), which has been used to model phonetic categorization (McMurray & Spivey, 2000). These types of models could

show classic online processing effects, while distinguishing anadrome pairs using fine-grained detail. In the meantime, these results challenge current models and the assumption that words are represented in terms of phoneme order, and they also challenge the field to develop creative alternatives to representing temporal order in models of word recognition.

Author Note We thank Jennifer Cole, Gary Dell, Simon Fischer-Baum, and Deborah Gagnon for insightful comments, and Matt Goldrick and James McQueen for helpful comments on earlier drafts. We also thank Dan McEchron for assistance with the data collection and processing. This research was supported by a Beckman Institute postdoctoral fellowship to J.C.T., NIH Grant DC008089 to B.M., and NIH Grant HD044458 to Gary Dell.

Appendix

Table 2 List of words used in the experiment

Base Word	Anadrome	Cohort	Unrelated	Feature Differences
tack	cat	tap	mill	1
pit	tip	pig	wine	1
puck	cup	putt	mail	1
pack	cap	pan	chess	1
reed	deer	reel	phone	1
face	safe	faith	lime	1
tug	gut	tub	mice	2
side	dice	sign	witch	2
lute	tool	loon	judge	2
mug	gum	mud	fish	2
mad	dam	map	shoes	2
sick	kiss	sip	jam	2
lip	pill	lid	cow	3
leak	keel	leash	gem	3
leap	peel	leaf	moon	3
sub	bus	sun	well	3

The final column indicates the number of feature differences (voicing, place, and manner) in the initial consonant between the anadrome and target.

References

- 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. *Journal of Memory and Language*, 38, 419–439. doi:10.1006/jmla.1997.2558
- Bates, D., & Sarkar, D. (2011). lme4: Linear mixed-effects models using Eigen and Eigen++ [R software package]. Retrieved from CRAN.
- Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of Acoustic Society of America*, 66, 1001–1017.
- Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, 18, 225–241. doi:10.1016/S0022-5371(79)90136-1
- Connine, C. M., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57, 483–501.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language & Cognitive Processes*, 16, 507–534. doi:10.1080/01690960143000074
- Frauenfelder, U., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language & Cognitive Processes*, 16, 563–607.
- Gaskell, M. G., & Marslen-Wilson, W. (1997). Integrating form and meaning: A distributed model of speech perception. *Language & Cognitive Processes*, 12, 613–656.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279. doi:10.1037/0033-295X.105.2.251
- Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Sciences*, 8, 58–59. doi:10.1016/j.tics.2003.11.006
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423–445.
- Guerrara, C., & Forster, K. (2008). Masked form priming with extreme transposition. *Language & Cognitive Processes*, 23, 117–142.
- Hannagan, T., Dupoux, E., & Christophe, A. (2011). Holographic string encoding. *Cognitive Science*, 35, 79–118. doi:10.1111/j.1551-6709.2010.01149.x
- Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62, 615–625. doi:10.3758/BF03212113
- 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.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. doi:10.1016/0010-0285(86)90015-0
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 306–321. doi:10.1037/0278-7393.31.2.306
- McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60, 1–39. doi:10.1016/j.cogpsych.2009.06.003
- McMurray, B., & Spivey, M. J. (2000). The categorical perception of consonants: The interaction of learning and processing. *Proceedings of the Chicago Linguistics Society*, 34, 205–220.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86, B33–B42.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–395. doi:10.1037/0033-295X.115.2.357
- Perea, M., & Lupker, S. J. (2003). Does *jugde* activate *COURT*? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition*, 31, 829–841. doi:10.3758/BF03196438
- Perea, M., & Lupker, S. J. (2004). Can *CANISO* activate *CASINO*? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51, 231–246. doi:10.1016/j.jml.2004.05.005
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298. doi:10.1016/j.jml.2007.04.001
- Spivey, M. J. (2007). *The continuity of mind*. New York, NY: Oxford University Press.
- 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. doi:10.1126/science.7777863