



Brief article

The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech

Hadas Shintel *, Howard C. Nusbaum

Department of Psychology and Center for Cognitive and Social Neuroscience, The University of Chicago, Beecher 102, 5848 South University Avenue, Chicago, IL 60637, USA

Received 3 August 2006; accepted 15 November 2006

Abstract

Language is generally viewed as conveying information through symbols whose form is arbitrarily related to their meaning. This arbitrary relation is often assumed to also characterize the mental representations underlying language comprehension. We explore the idea that visuo-spatial information can be analogically conveyed through acoustic properties of speech and that such information is integrated into an analog perceptual representation as a natural part of comprehension. Listeners heard sentences describing objects, spoken at varying speaking rates. After each sentence, participants saw a picture of an object and judged whether it had been mentioned in the sentence. Participants were faster to recognize the object when motion implied by speaking rate matched the motion implied by the picture. Results suggest that visuo-spatial referential information can be analogically conveyed and represented.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Spoken language comprehension; Perceptual representations, Prosody

* Corresponding author.

E-mail address: hadas@uchicago.edu (H. Shintel).

1. Introduction

Language is generally viewed as a symbolic system in which semantic-referential information is conveyed through arbitrary discrete symbols – there is no inherent relation between form and meaning. In fact, this arbitrary relation between form and meaning is commonly accepted as an essential characteristic of linguistic signs (Hockett, 1960; Saussure, 1959), in contrast to iconic signs whose form corresponds in some way to what they represent (cf. Peirce, 1932). In contrast to words, several accounts have suggested that prosodic properties of speech do constitute motivated signs that exhibit non-arbitrary form–meaning relations (Bolinger, 1964, 1985; Gussenhoven, 2002; Ohala, 1994). However, the role of prosody has been viewed as limited to conveying information about the message or about the speaker, rather than directly conveying information about external referents. For example prosody has been shown to convey information about the syntactic structure of the message or about the discourse status of the information it conveys (e.g. Birch & Clifton, 1995; Snedeker & Trueswell, 2003), as well as information about the speaker’s emotion or attitude (e.g. Banse & Scherer, 1996; Bryant & Fox Tree, 2002). But prosodic information has been viewed as affecting referential interpretation only in so far as it allows listeners to infer the intended referent given information about discourse structure or speaker’s attitude.

However, manipulation of non-symbolic continuous acoustic properties of speech has the potential of directly conveying semantic-referential information. Research on non-speech sounds has shown that people perceive cross-modal correspondences between auditory and visual sensory attributes, for example between pitch and various visuo-spatial properties such as vertical location, size, and brightness (e.g. Marks, 1987) and moreover, that such cross-modal correspondences influence perceptual processing. For example classification of the vertical position of a visual target was facilitated by a congruent-frequency sound (high position-high frequency) and impaired by an incongruent-frequency sound (Bernstein & Edelstein, 1971; Melara & O’Brien, 1987), suggesting a cross-modal association between pitch height and vertical location. A similar congruency effect was found for pitch and the spoken or written words *HIGH* and *LOW* (Melara & Marks, 1990).

Although this issue has rarely been investigated, cross-modal correspondences may be functional in everyday communication. Speakers can convey referential information by mapping visual information onto acoustic–auditory properties of speech, capitalizing on existing auditory–visual mappings. For example Shintel, Nusbaum, and Okrent (2006) showed that when speakers were instructed to describe an object’s direction of motion by saying either *it’s going up* or *it’s going down*, they spontaneously raised and lowered the fundamental frequency of their voice (the acoustic correlate of pitch), mapping fundamental frequency to described direction of motion; when instructed to describe the horizontal direction of motion (left vs. right) of a fast- or a slow-moving object, speakers spontaneously varied their speaking rate, mapping articulation speed to visual speed of object motion. Furthermore, listeners could interpret information about objects’ speed conveyed exclusively through prosody; listeners were reliably better than chance at classifying speed of

motion (fast vs. slow) from sentences describing only the object's direction of motion. Classification accuracy was significantly correlated with utterance duration (positive accuracy-duration correlation for utterances describing slow-moving objects, negative correlation for utterances describing fast-moving objects), suggesting duration was the basis for classification. These findings suggest that such analog acoustic expression is a natural part of spoken-communication; rather than relying exclusively on arbitrary linguistic symbols, non-arbitrary analog signs can directly provide independent referential information.

While the assumption regarding the arbitrary nature of linguistic signs concerns external signs (such as words), it finds its counterpart in the critical assumption in many theories in cognitive science (e.g. Fodor, 1975; Pylyshyn, 1986) about the structure of the mental representations underlying language comprehension (or cognition in general). According to this assumption, the structure of external linguistic signs parallels the language-like structure of the mental representations underlying the use of these signs. Such mental representations are generally considered to be abstract symbols whose form is arbitrarily related to what they represent. However, recent research suggests that language comprehension involves perceptual-motor representations that are grounded in actual perceptual-motor experience and analogically related to their referents (Barsalou, 1999; Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000; Zwaan & Madden, 2004). Unlike amodal abstract symbols, perceptual symbols are modal, that is represented in the same perceptual system that produced them, and analogical, that is the structure of the representation corresponds to the structure of the represented object or of the perceptual state of perceiving the object (Barsalou, 1999). Thus, in contrast to amodal representations that are not directly connected to their real-world referents (see Harnad, 1990), analog modal representations are grounded in actual processes of sensorimotor interaction with real-world referents.

Several findings have shown that language comprehension routinely involves activation of perceptual information about objects' shape, orientation, and direction, that is implied by sentences (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; Zwaan, Madden, Yaxley, & Aveyard, 2004). Zwaan et al. (2002) showed that participants were faster to verify that a drawing represents an object that had been mentioned in a sentence when the object's shape in the drawing matched the shape implied by the sentence compared to when there was a mismatch between them. For example participants were faster to verify that a drawing of an eagle with outstretched wings represents a mentioned object following the sentence "The ranger saw the eagle in the sky" than after the sentence "The ranger saw the eagle in the nest". This pattern of results is not predicted by accounts that claim that sentence meaning is represented by a propositional representation that does not refer to perceptual shape. Importantly, these results suggest that comprehension involved perceptual representations even though participants' task did not require the use of such information.

If non-propositional analog representations are indeed involved in language comprehension, analog acoustic expression may provide a particularly apt signal for such a form of representation. Unlike words, in this case the external signal itself is analog

and non-arbitrary. By analogically mapping variation in the referential domain onto variation in speech, analog expression may provide a kind of grounded representation and a non-arbitrary form–meaning mapping that may facilitate comprehension.

The present experiment investigated whether referential information conveyed exclusively through analog acoustic expression, specifically motion information, is integrated into a perceptual representation of the referent object. Previous research (Shintel et al., 2006) suggests that speaking rate can convey information about objects' speed of motion, even when the propositional content of the utterance involves no reference to speed. However, that study used an explicit speed classification task which required listeners to go beyond the propositional content and may have forced them to rely on acoustic properties of speech that they do not typically attend to or use as a source of referential information. Listeners may not routinely use this information in comprehension when they are not faced with a decision that depends on it. If, on the other hand, information conveyed through analog variation of acoustic properties of speech is interpreted naturally during comprehension, listeners may integrate it into their representation of the object. For example, listeners may be more likely to represent the object as moving after hearing a sentence spoken at a fast speaking rate, even if the propositional content of the sentence does not refer to movement. Furthermore, listeners may represent analogically conveyed information in a homologous form that can be integrated into an analog perceptual representation of the object. For example, the perceptual representation of a fast-spoken sentence describing an object may correspond to the visual experience of seeing the object in motion.

To evaluate this question, we used a task modelled after the paradigm used by Zwaan et al. (2002) in which participants had to determine whether a picture represents an object that had been mentioned in a previous sentence. The task was merely to determine if the picture represents an object of the same category as the object mentioned in the sentence. In contrast to the classification task used in our previous research, in which listeners judged the described object's speed of motion, the present task did not require the use of motion information. Listeners heard a sentence describing an object, spoken at a fast or a slow rate. The propositional content of the sentence did not refer to, or imply any motion information. Following each sentence, listeners saw a picture of the object mentioned as the sentence subject. Some participants saw a picture of the object in motion, while others saw a picture of the object at rest (see Fig. 1). Studies have shown that static images of objects in motion can imply object motion (Freyd, 1983; Kourtzi & Kanwisher, 2000). Thus the picture either implied or did not imply that the object is moving. If fast speech rate can imply object motion, and if listeners understand the referent of a sentence by integrating information conveyed through analog acoustic expression into a perceptual representation of the propositionally described object, then participants should be faster verifying that the depicted object had been mentioned in the sentence when motion implied in the picture is congruent with motion implied in speech rate (fast speech rate – moving object) compared to the incongruent condition (slow speech rate – moving object).



Fig. 1. Example of picture stimuli used in the experiment for the sentence “The horse is brown”. The “rest” picture depicts a standing horse; the “Motion” picture depicts a running horse.

2. Method

2.1. Participants

Thirty four University of Chicago students participated in the study. All participants had native fluency in English and no reported history of speech or hearing disorders. Participants were paid for their participation.

2.2. Materials

Test stimuli included 16 sentences that described different objects. None of the sentences referred to movement or implied that the described object was moving or not moving. Each sentence was paired with two pictures (never displayed to the same participant) depicting the object mentioned as the sentence subject. In all test stimuli the displayed object matched the description in the sentence. One of the pictures depicted the object in motion; the other picture depicted the same object at rest. In addition, 16 filler sentences were paired with 16 additional pictures. Filler pictures never depicted an object mentioned in the corresponding sentence (therefore conveying no information about the mentioned object’s motion). Sentences were produced by a female speaker. Each test sentence was recorded twice: once spoken at a “fast” speech rate and once spoken at a “slow” speech rate (mean WPM 282 and 193 for the fast- and the slow-spoken sentences, respectively, mean syllables per word = 1.3). The speaker produced the test sentences while watching a fast- or a slow-moving time-bar on the computer and tried to match the speed of her speech to the speed of motion of the bar. Prior to recording the stimulus sentences, the speaker was asked to speak a select sample of the sentences at different speech rates. Time-bars duration was determined based on the duration of these sentences. Filler sentences were produced at the speaker’s natural speaking rate, spontaneously varying across different sentences (the speaker’s natural speaking rate was somewhat closer to the slow speech, mean WPM 212). For test and filler sentences, other acoustic properties

such as amplitude and fundamental frequency varied with the way the speaker naturally produced them. Sentences were recorded using a SHURE SM94 microphone onto digital audiotape and digitized at a 44.1 kHz sampling rate with 16-bit resolution. Utterances were edited into separate sound files beginning with the onset (first glottal pulse) of each sentence.

2.3. Design and procedure

Speech Rate (fast vs. slow) and Picture (motion vs. rest) were manipulated within subjects. Each participant was presented with 16 test items, four in each Speech Rate \times Picture combination. We created four lists that counterbalanced items across subjects. Additionally each participant was presented with 16 filler sentences. Sentences were presented in random order. As “motion” and “rest” pictures differed substantially, response times cannot be compared across the two Picture conditions. To compare object recognition times for the two picture types, six additional participants completed a version of the task in which the pictures followed a written version of the test sentences. Results showed reliably shorter reaction times for “rest”, compared to “motion”, pictures (609 and 695 ms, respectively, $t(5) = 2.58$, $p < .05$). This difference may be due to visual differences between the pictures or to “rest” pictures being the more typical representations of the objects. Thus, the critical comparisons concern the effect of Speech Rate within each Picture condition.

Participants sat in front of a computer and heard the sentences through headphones. Each sentence was followed by a fixation point in the middle of the screen for 250 ms. Following the fixation, participants saw a picture of an object and had to determine whether it was mentioned in the preceding sentence and respond with their dominant hand by pressing keys marked “YES” and “NO”. Participants were instructed to respond “YES” if the depicted object belonged to the same category as the object in the sentence (e.g. if the sentence mentions a horse and the picture displays a horse). This was done in order to emphasize that the task is a categorization task that does not require the use of motion information or properties other than its category membership.

3. Results and discussion

Response times greater than 2.5 standard deviations above the subject’s mean were excluded from the analyses. Given the small number of test trials, if two trials or more were affected by the trimming procedure ($>10\%$), data for the subject were excluded from the analysis. This resulted in excluding data from two subjects. Within the subjects who were included in the analysis, the trimming procedure affected a total of 9 (mean RT 1684 ms) out of 512 trials ($<2\%$ of the trials).

Analysis of response accuracy showed responses were almost always correct. Subjects made incorrect responses (responding ‘No’ when the picture showed a mentioned object) only on six trials: two congruent trials (one fast speech/‘motion’ picture and one slow speech/‘rest’ picture) and four incongruent trials (all slow

speech/‘motion’ picture) out of 512 (<2% of the trials). Accuracy scores did not differ reliably between conditions (all $p > .1$, n.s.). These trials were not included in the response time analysis.

Analysis of response times showed that subjects were faster to respond when the motion conveyed through analog acoustic expression matched the motion implied in the picture (fast speech/“motion” picture and slow speech/“rest” picture: mean 624.33 ms, SEM 20.1) compared to the condition in which analog acoustic information did not match the picture (fast speech/“rest” picture and slow speech/“motion” picture: mean 661.05 ms, SEM 26.6). A repeated measures ANOVA with Speech Rate (fast vs. slow) and Picture (motion vs. rest) as within-subjects factors revealed a significant Speech Rate by Picture interaction ($F(1,31) = 5.369$, $MSE = 43,153.7$, $p < .03$). The main effects of Speech Rate and of Picture were not significant ($p > .2$).¹

A simple effects analysis of the effect of speech rate on listeners’ response latencies for each picture type showed a reliable effect of speech rate on recognition of “motion” pictures; listeners responded faster to “motion” pictures when these were preceded by congruent fast speech compared to incongruent slow speech (621 and 681 ms, respectively, $t(31) = 2.68$, $p < .01$). There was no reliable effect of speech rate on “rest” pictures (628 ms for slow speech and 641 ms for fast speech, $t(31) = .76$, $p > .2$), although the pattern was in the same direction as the congruency effect for “motion” pictures. This pattern of results suggests that the slightly more unusual fast speech rate provides a benefit for recognizing the more atypical, or less expected, object pictures.² However, slow speech rate does not provide a reliable advantage for recognizing the more typical object representations. It may be that speech rate needs to deviate more from an average speaker’s typical speech rate to affect listeners’ expectations about objects, and consequently their mental representations of objects. Our speaker’s natural rate of speech for the filler sentences was closer to the slow sentences than to the fast sentences. It is possible that given the similarity of the slow speech rate to the speaker’s typical speech rate, it did not reliably affect listeners’ expectations about objects. Furthermore, it is possible that listeners expect slower speech rate that is closer to a standard of ‘clear speech’ in the context of a psychology experiment. Finally, even in contexts in which a slower speech rate is relatively distinct, and thus may be more informative for listeners, the mapping between speech rate and implied object motion is more ambiguous in the case of slow speech. For example, slow speech may be mapped to slow motion, rather than to non-motion; a distinction between fast- and slow-moving objects is difficult to recreate with static images. Further research is needed to examine these alternatives.

¹ Due to the small number of items, the Speech Rate by Picture interaction was not reliable in the item analysis ($F(1, 15) = 2.01$, $MSE = 16051$, $p = .17$), however results showed the same pattern. Main effects of Speech Rate and of Picture were not significant (both effects $F < 1$, $p > .4$). Effect size for Speech Rate within each of the Picture conditions using Cohen’s d adjusted for repeated measures (Dunlap, Cortina, Vaslow, & Burke, 1996) was .442 for ‘motion’ pictures, and .162 for the ‘rest’ pictures.

² Object recognition times were longer for “motion” compared to “rest” pictures, see Section 2.2.

Results show that listeners are sensitive to information conveyed exclusively through analog acoustic expression and integrate it into their representation of the referent object as a natural part of comprehension. Listeners spontaneously used this information even when the task did not explicitly or implicitly require its use. Indeed, attending to analogically conveyed motion information did not confer any performance benefit for several reasons. First, half of the pictures depicted unmentioned objects. In these cases, speaking rate would be irrelevant to the decision. Second, pictures depicting mentioned objects were just as likely to be incongruent with the analog acoustic information as they were to be congruent with it. This suggests that listeners use this information as a natural part of comprehension, rather than as a strategic decision process. Moreover, all pictures depicted objects that clearly matched the verbal description in the sentence (e.g. the sentence “The horse is brown” was always followed by a picture depicting a brown horse). Finally, given the small number of congruent trials (four fast-speech/moving-object trials and four slow-speech/resting-object trials, or 25% of all trials), it is unlikely that participants noticed a relation between speech rate and the picture, making it unlikely that they could have intentionally used this information to develop expectations about the picture.

The relation between speech rate and object motion in comprehension can be explained by several possible underlying processes. First, listeners may rely on a cross-modal audio–visual similarity between rate of visual motion and rate of articulation. The relation between fast speech and object motion may thus be similar to the relation between high pitch and high vertical position. Second, this relation may be based on a learned association between faster speech rate and object motion. Speakers may speak faster when describing dynamic states of affairs (which frequently involve some sort of motion) compared to static situations. Listeners may come to associate a faster speech rate with motion as a result of this co-occurrence. Third, a faster speech rate may be attributed to urgency on the part of the speaker; speaker’s urgency may imply a more dynamic situation. Our previous research (Shintel et al., 2006) suggests that speakers vary their speech rate when they are describing fast motion even when such variation is not required by the situation; participants spoke faster when describing fast-moving dots even though the duration of the display was the same and was significantly longer than the average duration of the descriptions. Thus variation in speech rate cannot be explained merely as a result of task demands or of an objectively time-sensitive situation. However, it is possible that listeners interpret faster speech rate as indicative of urgency. Finally, it should be noted that these explanations need not be mutually-exclusive.

Given that listeners spontaneously use information conveyed by speech rate, the performance advantage observed in the congruent condition (when acoustically conveyed motion matched the motion implied in the picture) suggests that understanding the sentence and the picture may depend on similar representations. A better match between these representations may facilitate recognition.

The view that language comprehension involves analog perceptual representations offers an explanation for our results. If listeners construct a perceptual representation of the verbally described object and integrate analog acoustic

information into that representation, the congruent condition should offer a closer match to the visual representation constructed while seeing the pictures. Although there will still be discrepancies between the sentence-generated representation and the picture-generated representation (the direction of motion, background, etc.), the closer match may facilitate recognition.

Of course, it is possible that listeners represent analog acoustic information in an abstract proposition rather than perceptually. Listeners would have to convert analog acoustic information into a propositional or featural representation, perhaps by augmenting the sententially-derived proposition with a property such as [MOVING]. If pictures are also represented in discrete propositional form, the closer match between these representations could facilitate performance.

Although we cannot rule out a purely propositional account, our results seem more consistent with similar studies that have been interpreted as suggesting that language comprehension involves perceptual representations (see Zwaan & Madden, 2004). In addition, several studies support the idea of perceptually dynamic mental representations (see Freyd, 1987), and such dynamic representations may be involved in language comprehension (Zwaan et al., 2004). Although the present study does not provide evidence for dynamic mental representations, it raises the possibility that dynamic information can be analogically conveyed through time-changing acoustic properties of speech, even when the propositional content does not imply such information. Further work is needed to evaluate the exact form of the representations underlying the findings of the present study.

Our results suggest that spoken sentences can contain information that goes beyond the words and the propositional structure. Acoustic properties of speech, like the gestures accompanying speech (Goldin-Meadow, 1999; McNeill, 1992), can convey analogical information about objects. Prosody functions not just to signal speaker's internal states, but must be understood scientifically as a source of referential information that can be varied independent of the lexical-propositional content of an utterance.

Acknowledgments

We thank Rachel Hilbert and Ashley Swanson for their help with the experiment. We thank Rolf Zwaan and three anonymous reviewers for their helpful comments on the paper. The support of the Center for Cognitive and Social Neuroscience at The University of Chicago is gratefully acknowledged.

References

- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology*, 70(3), 614–636.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, 22, 577–660.
- Bernstein, I., & Edelman, B. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, 87, 241–247.

- Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech*, 38, 365–391.
- Bolinger, D. L. (1964). Intonation across languages. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language phonology* (Vol. 2). Stanford, CA: Stanford University Press.
- Bolinger, D. (1985). The inherent iconicism of intonation. In J. Haiman (Ed.), *Natural syntax: iconicity and erosion*. Cambridge, UK: Cambridge University Press.
- Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor & Symbol*, 17(2), 99–117.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170–177.
- Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell.
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception and Psychophysics*, 33, 575–581.
- Freyd, J. J. (1987). Dynamic mental representation. *Psychological Review*, 94, 427–438.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychological Bulletin & Review*, 9, 558–565.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Science*, 3, 419–429.
- Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence: ProSig and Université de Provence Laboratoire Parole et Langage.
- Harnad, S. (1990). The symbol grounding problem. *Physica, D* 42, 335–346.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3), 88–96.
- Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, 12, 48–55.
- Marks, L. E. (1987). On cross-modal similarity: auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, 13(3), 384–394.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.
- Melara, R., & Marks, L. (1990). Processes underlying dimensional interactions: correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, 18, 477–495.
- Melara, R., & O'Brien, T. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116, 323–336.
- Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism*. Cambridge, UK: Cambridge University Press.
- Peirce, C. S. (1932). Division of signs. In C. Hartshorne & P. Weiss (Eds.), *Collected papers of C.S. Peirce* (Vol. 2). Cambridge, MA: Harvard University Press.
- Pylyshyn, Z. W. (1986). *Computation and cognition: toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Saussure, F. de. (1959). *Course in general linguistics*. New York and London: McGraw-Hill.
- Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech. *Journal of Memory and Language*, 55, 167–177.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12(2), 153–156.
- Zwaan, R. A., & Madden, C. J. (2004). In D. Pecher & R. A. Zwaan (Eds.), *The grounding of cognition: the role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.
- Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: dynamic representations in language comprehension. *Cognitive Science*, 28, 611–619.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13, 168–171.