Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 105 (2007) 681–690
COGNITION
www.elsevier.com/locate/COGNIT

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.

Keywords: Spoken language comprehension; Perceptual representations, Prosody

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

0010-0277/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2006.11.005

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